

Qy 2020 TCATGTTTAAAAA 2079
 Db 96 GAACCTTTAAAAA 37
 Qy 2080 AAAAAA 2106
 Db 36 AAAAAA 10

RESULT 4
 LOCUS BM884782
 DEFINITION rcl5a03.y1 Meloidogyne hapla egg pAMP1 v1 Meloidogyne hapla cDNA
 5', mRNA sequence.
 ACCESSION BM884782
 VERSION BM884782.1 GI:19268526
 KEYWORDS EST.
 SOURCE Meloidogyne hapla
 ORGANISM Meloidogyne hapla

REFERENCE 1 (bases 1 to 336)
 AUTHORS McCarter, J., Clifton, S., Chiapelli, B., Pape, D., Martin, J., Wylie, T., Darter, M., Marra, M., Hillier, L., Kucaba, T., Theising, B., Bowers, F., Gibbons, M., Ritter, E., Bennett, J., Franklin, C., Tsagarelis, V., Ronko, I., Kennedy, S., Maguire, L., Beck, C., Underwood, K., Steptoe, M., Allen, M., Person, B., Swaller, T., Harvey, N., Schurk, R., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilton, R.

TITLE The Washington Univ. Nematode EST Project, 1999
 JOURNAL Unpublished
 COMMENT Contact: McCarter JP
 The Washington Univ. Nematode EST Project, 1999
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 The library was constructed by Claire Murphy and Dr. James McCarter at Washington University, St. Louis. DNA Sequencing by: Washington University Genome Sequencing Center
 Seq primer: -40RP from Gibco
 High quality sequence stop: 302.

FEATURES
 source
 1. 336
 /organism="Meloidogyne hapla"
 /mol_type="mRNA"
 /db_xref="taxon:6305"
 /dev_stage="parasitic adult females"
 /lab_host="DH10B"
 /clone_lib="Meloidogyne hapla egg pAMP1 v1"
 /note="Vector: pAMP1 (Gibco); Site.1: NotI; Site.2: SalI;
 The library was constructed by Claire Murphy and Dr. James McCarter at Washington University, St. Louis. The cDNA was made by using Dynabead oligo-dT priming (Dyna). PCR based library using a modified protocol from the SMART PCR cDNA Synthesis kit from Clontech. Directionally cloned into the UDG sites of pAMP1."

BASE COUNT 207 a 21 c 31 g 77 t
 ORIGIN

Query Match 4.5%; Score 95.4; DB 12; Length 336;
 Best Local Similarity 72.8%; Pred. No. 9.8;
 Matches 123; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

Qy 1938 ATGTGTGATGTCAGCAGCGGTTGAAAGATTTTGTCTATCTAGTACGACCTGCTATTGTC 1997
 Db 154 ATGTTTATTTTATATATTTTAAAAAATTCATGTTTCACAAAATTTCTGTTTGTGA 213

Qy 1998 TGAATAGCTTATGCGACATCATGCTTTTAAAAA 2057
 Db 214 TAAAAATTTTGCCTTTTGA 273

Qy 2058 AAAAAA 2106
 Db 274 AAAAAA 322

RESULT 5
 LOCUS AI392214/c
 DEFINITION NCSP19F7 Subtracted Perithecial Neurospora crassa cDNA clone SP19
 3' similar to mnt1 protein homolog - thiamine biosynthesis enzyme,
 mRNA sequence.
 ACCESSION AI392214
 VERSION AI392214.1 GI:4220021
 KEYWORDS EST.
 SOURCE Neurospora crassa
 ORGANISM Neurospora crassa

REFERENCE 1 (bases 1 to 480)
 AUTHORS Nelson, M.A., Kang, S., Braun, E.L., Crawford, M.E., Dolan, P.L., Leonard, P.M., Mitchell, J., Armijo, A.M., Bean, L., Blueyes, E., Cushing, T., Errett, A., Fleharty, M., Gorman, M., Judson, K., Miller, R., Ortega, J., Pavlova, I., Perea, J., Todisco, S., Trujillo, R., Valentine, J., Wells, A., Werner-Washburne, M., Yazzie, S. and Natvig, D.O.

TITLE Expressed sequences from conidial, mycelial, and sexual stages of Neurospora crassa
 JOURNAL Fungal Genet. Biol. 21, 348-363 (1997)
 MEDLINE 97435549
 PUBMED 9290248
 COMMENT Contact: Natvig, D.O./Nelson, M.A.
 Department of Biology
 University of New Mexico
 Castetter Hall, Albuquerque, NM 87131, USA
 Tel: 505 277 3411
 Fax: 505 277 0304
 Email: ngp@biology.unm.edu

FEATURES
 source
 1. 480
 /organism="Neurospora crassa"
 /mol_type="mRNA"
 /strain="fl a (FGSC 4347)"
 /db_xref="taxon:5141"
 /clone="SP19"
 /sex="Mating type a (fluffy), fertilized"
 /tissue_type="Perithecia"
 /dev_stage="Fruiting Body"
 /lab_host="E. coli"
 /clone_lib="Subtracted Perithecial"
 /note="Vector: pBluescript SK (-); Site.1: EcoRI; Site.2: XhoI; mRNA isolated from 5 day old perithecia (fruiting bodies) of the fluffy strain fl a (Mating type a), type A). cDNA directionally cloned into pBluescript SK(-) using the Uni-ZAP XR vector system (Stratagene, La Jolla, CA). Previously identified highly expressed clones were subtracted from this library."

BASE COUNT 128 a 137 c 48 g 166 t 1 others
 ORIGIN

Query Match 4.5%; Score 95.4; DB 9; Length 480;
 Best Local Similarity 67.2%; Pred. No. 7.9;
 Matches 135; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

Qy 1906 ACTTCAAGTTGAATGCATCTCTGGTCCCTACCATGTTGATGTACATAGCGGCTTGA 1965
 Db 209 AGTTCAGATGACAGTTTCTGTTGGATGATGGGCGCTCAGGGGTAGTAGGGAAG 150

Qy 1966 GATTTGCTGATTACTGAACCTGTACTTCTGCTGAATAGTTATGGCATATGATTCATGT 2025
 Db 149 GGTTGGGTGAGTAAAGTTGTTTTCAGTTTAAATGGGTACGTTTTTAAAAA 90

Qy 2026 TTA 2085

RESULT 10
BI746603
LOCUS
DEFINITION
BIF746603 165 bp mRNA linear EST 25-SEP-2001 rn29d01.yl Meloidogyne arenaria egg PAMP1 vI Chiapelli McCarter Meloidogyne arenaria cDNA 5', mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE / ORGANISM
Meloidogyne arenaria Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina; Tylenchoidea; Heteroderidae; Meloidogyninae; Meloidogyne. 1 (bases 1 to 165)
REFERENCE
AUTHORS
McCart,J., Clifton,S., Chlapelli,B., Pope,D., Martin,J., Wylie,T., Dante,M., Marra,M., Hillier,L., Kucaba,T., Theising,B., Bowers,Y., Gibbons,M., Rittter,E., Bennett,J., Franklin,C., Tsagarelshvili,R., Ronko,I., Kennedy,S., Maguire,L., Beck,C., Underwood,K., Steptoe, M., Allen,M., Person,B., Swaller,T., Harvey,N., Schurk,R., Kohn,S., Shin,T., Jackson,Y., Cardenas.M., McCann,R., Waterston,R. and Wilson,R.
TITLE
JOURNAL
COMMENT
The Washington Univ. Nematode EST Project, 1999 Unpublished Contact: McCarter JP The Washington Univ. Nematode EST Project, 1999 Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu The library was constructed by Brandi Chiapelli and Dr. James McCarter (bchiapel@watson.wustl.edu & jmccarte@watson.wustl.edu) at Washington University, St. Louis. DNA Sequencing by: Washington University Genome Sequencing Center St. Louis. Nematodes were provided by Dr. David Bird and Daniel Snyder of North Carolina State University. Seq primer: -4ORP from Gibco High quality sequence stop: 164.

```

FEATURES             High quality sequence stop: 164.
source              Location/Qualifiers
1. .165
   /organism="Meloiodogyne arenaria"
   /mol_type="mRNA"
   /db_xref="taxon:6304"
   /dev_stage="egg"
   /lab_host="DH10B"
   /clone_lib="Meloiodogyne arenaria egg pAMP1 v1 Chiapelli
McCarter"
   /notes=Vector: pAMP1 (Gibco); Site_1: NotI; Site_2: SalI;
The library was constructed by Brandi Chiapelli and Dr.
James McCarter at Washington University, St. Louis. The
cDNA was made by using Dynabead oligo-dT priming (Dynal). The
PCR based library using a modified protocol from the
SMART PCR cDNA Synthesis Kit from Clontech. Directionally
cloned into the UDG sites of pAMP1. Nematodes were
provided by Dr. David Bird and Daniel Snyder of North
Carolina State University."
BASE COUNT          117 a      8 c      6 g      34 t
ORIGIN
Query Match         4.5%; Score 94.2; DB 12; Length 165;
Best Local Similarity 77.6%; Pred. No. 19;
Matches 114; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
QY    1960 TTGAAGATTTTCGTATTACTGAACCTGACTTTTGTCGAANTAGTATGGCACTATGAT 2019
       || ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     8 TTTAAGATTTTTTTTAAACTTCGCCTTAATTTATTTGAANTAAAACGTCCCTTGATT 67
       || ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY    2020 TCATGTTTAAAAA AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2079
       | ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     68 AAAA AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 127
       ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY    2080 AAAAAA AAAAAAAAAAAAAAAAAAAAAA 2106
       ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |

```

[illegible]

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 24, 2003, 01:59:13 ; Search time 133 Seconds

(without alignments)

6989.119 Million cell updates/sec

Title: US-09-856-327-1

Perfect score: 2106

Sequence: 1 atcagcagctctctctctca.....aaaaaaaaaaaaaaaaaaaaa 2106

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_NA:*

1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*

2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*

3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*

4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*

5: /cgn2_6/ptodata/2/ina/PTUS_COMB.seq.*

6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	212.4	10.1	1869	3	US-09-305-381-1
2	210.2	10.0	1869	1	US-08-734-925-1
3	152.6	7.2	1701	3	US-09-023-731-2
4	91.6	4.3	2323	4	US-09-149-476-24
5	88.8	4.2	1733	3	US-09-073-569-1
6	88.4	4.2	1091	4	US-09-328-965-1
7	88	4.2	1248	4	US-09-489-847-101
8	88	4.2	2246	4	US-09-363-708-3
9	88	4.2	2246	4	US-09-083-587-3
10	87.2	4.1	2434	4	US-09-489-847-67
11	87	4.1	1114	4	US-09-152-060-41
12	87	4.1	2239	4	US-09-196-390-1
13	86.6	4.1	2671	6	5168051-9
14	85.6	4.1	1052	4	US-09-489-847-23
15	85	4.0	1582	3	US-08-545-196B-10
16	85	4.0	1582	3	US-08-545-196B-12
17	84.8	4.0	578	4	US-09-602-877A-95
18	84.8	4.0	3060	4	US-09-996-243-504
19	84.8	4.0	3715	4	US-09-234-245-1
20	84.8	4.0	7724	4	US-08-486-049-1
21	84.6	4.0	1037	4	US-09-489-847-112
22	84.6	4.0	1696	4	US-09-835-811-1
23	84.4	4.0	1249	4	US-09-461-325-128
24	84.4	4.0	1641	1	US-08-300-903A-8
25	84.4	4.0	1641	4	US-08-988-197-8
26	84.4	4.0	3410	3	US-09-020-956-110
27	84.4	4.0	3410	3	US-09-030-607-110

28	84.4	4.0	3410	4	US-09-439-313-110	Sequence 110, App
29	84.4	4.0	3410	4	US-09-352-616A-110	Sequence 110, App
30	84.4	4.0	3410	4	US-09-602-877A-100	Sequence 100, App
31	84.4	4.0	3410	4	US-09-232-149A-110	Sequence 110, App
32	84	4.0	441	4	US-09-601-537-10	Sequence 10, Appl
33	84	4.0	4121	4	US-09-601-537-9	Sequence 9, Appl
34	83.6	4.0	1474	3	US-08-821-994-64	Sequence 64, Appl
35	83.6	4.0	6409	4	US-09-967-908A-1	Sequence 1, Appl
c	83.4	4.0	1260	4	US-09-461-325-93	Sequence 93, Appl
37	83.4	4.0	1602	1	US-08-530-950-3	Sequence 3, Appl
38	83.4	4.0	1602	3	US-08-888-429A-3	Sequence 3, Appl
39	83.4	4.0	1602	3	US-09-149-879-3	Sequence 3, Appl
40	83.4	4.0	1602	4	US-09-057-009-3	Sequence 3, Appl
41	83.4	4.0	1895	4	US-09-444-336-7	Sequence 7, Appl
42	83.2	4.0	674	4	US-09-620-405B-465	Sequence 465, App
43	83.2	4.0	674	4	US-09-433-826B-465	Sequence 465, App
44	83.2	4.0	674	4	US-09-604-287A-465	Sequence 465, App
45	83.2	4.0	1361	4	US-09-489-847-64	Sequence 64, Appl

ALIGNMENTS

RESULT 1

US-09-305-381-1

; Sequence 1, Application US/09305381

; Patent No. 6146865

; GENERAL INFORMATION:

; APPLICANT: Christensen, Soren

; APPLICANT: Lassen, Soren Flensted

; APPLICANT: Schneider, Palle

; TITLE OF INVENTION: Nucleic Acids Encoding Polypeptides

; TITLE OF INVENTION: Having Pyranose Oxidase Activity

; FILE REFERENCE: 5571.200-US

; CURRENT APPLICATION NUMBER: US/09/305.381

; CURRENT FILING DATE: 1999-05-05

; EARLIER APPLICATION NUMBER: 60/088,724

; EARLIER FILING DATE: 1998-06-10

; EARLIER APPLICATION NUMBER: PA 1998 00774

; EARLIER FILING DATE: 1998-06-08

; NUMBER OF SEQ ID NOS: 2

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 1

; LENGTH: 1869

; TYPE: DNA

; ORGANISM: Trameetes hirsuta

US-09-305-381-1

Query Match	10.1%	Score 212.4;	DB 3;	Length 1869;
Best Local Similarity	51.5%;	Pred. No. 6.7e-39;		
Matches	841;	Conservative	0;	Mismatches 656;
				Indels 135;
				Gaps 10;
Qy	269	GGCTACACACAGAAGATGAATCGAGTTCGAGTTCAGAAAGATATTGACCGCTTCGTCATGTGA 328		
Db	259	GGCTACACACAGAAGAACACCGTCGAGTACCAAGAAACATCGACAAATTCGTAATGTT 318		
Qy	329	ATCAAGGCGCTTACACAAAGTCTCTGTTCTCTGAGAACAGAGAGCTGCCTACACTT 388		
Db	319	ATACAAGGGCAACTTATGCCGCTCGTCCGTCACACAGCATGGTGTGACACCGCTA 378		
Qy	389	GATCCCGGAGCTGGAGCGCGCCCTTGAAGTTTACGCCATATCAACAGGTAAATTCCT 448		
Db	379	AGCCCGGCGTCATGGCAAGCTTTCGACG-----TTCCTCGCCGCAACGGGGCAATCCA 432		
Qy	449	CACCAGCGGAATTCGAGAACTTCTCTCGGAGGCGCGTCAACGGTGGAGTCGGCGCATG 508		
Db	433	GAGCAAGACCGCTGCGCAACCTTAGTGGCCAGCGGTCACCCGCTCGTCGGCGCATG 492		
Qy	509	AGTACCACCTGGAGCTCTCCACGCGACGGATTTCATCCACCCATGCGAAAGTCTCCCGGC 568		
Db	493	TCATACGACCTGGAGCTGCGGACCGCGGCTTCGAGAAAGCTGCAAG----- 537		
Qy	569	ATCGGCGCTCCGAAGCTCAGTAACGACCGCGCAGAGGACGACAAAGAGTGAACGAGCTT 628		

Best Local Similarity 50.48; Pred. No. 2.le-38;
Matches 857; Conservative 0; Mismatches 713; Indels 129; Gaps 9;

```
QY 200 GAGATCGAGTGTGATAGCTTCTACGCTGTTAAATCCGGAAGAAGCAATGCGAGTTCCC 259
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 190 GAGCTCTCGAAGCGGTTTCAAGGTCCGATGTCGACATCGGGGAAATGACTTGGC 249
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 260 TAGCTTCTCGCTACCAACAAGAAGAAATGAATCGAGTTCAGAAAGATATTGACCGCTTC 319
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 250 CTGAAGATCGGTGCCCAACAAGAAGAACACCGCTCGAATACCAAGAACAATTGACAAGTTT 309
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 320 GTCAATGTAATCAAGGAGCGCTTACAACAAGTCTCTGTTCCGTGTCAGAAACCAAGAGTG 379
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 310 GTGAACGTCATTCAGGCGCAATGATGCTGTCTGTTTCCGTTCCCGCTCAATACCCCTCGTATC 369
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 380 CTTACACTTGTATCCCGGAGCGCTGGAGCGCCGCCCTGGAAGTTCAGCCATATCGAACGGT 439
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 370 GACACGCTACGCCGAGCTTTGGCAAGCTTCATCG-----TTCCTCGTCCGCAATGGC 423
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 440 AAAAATCCTCACCAGCGGGAATTCGAGAATTTGTCGCGAGGCGGTAACGCGTGGAGTC 499
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 424 TCGAAACCCAGAGCAGGACCGCTTCGTAACCTCAGTGTGTCAGCGGTCACGCGTGTGCTC 483
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 500 GCGGGATGATGATCCCACTGAGCTGCTCCACGCCACGGATTATCCACCCATGGAAGT 559
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 484 GGAGGATGTCCACGCACTGGACATGCGGACACCGCGCTTTGA----- 527
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 560 CTCGCGGCATCGSCCTCGAAGCTCAGTAACGCCCGCAGAGGACGACAAAGATGG 619
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 528 --CGCGAGCAGCGCCGCTGCTGCTGAAGACGACGACGAGCTGACGAGCGGAGTGG 585
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 620 AACGAGCTTTATTCGAGGCGGAGCGTCTCATCGGACTTCCACCAAGGAATTCGACGAG 679
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 586 GACCGGTGTACACCAAGGCGGAGTCACTTCAAGACCGGGAGCGGACAGTTCGAAGGAG 645
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 680 TCAATTCGGACACACCTTGTCTCGCTCTTTGCAAGACGCGTACAGGATTCGTCACGT 739
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 646 TCGATCGGCACACACCTCGTGTCAACAGCTCGCGGAGGAATACAAAGTTCAGCGCGAC 705
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 740 ATCTTTTCGCCCTCTCCCGTTGGGATGCCACCGTTGAAGAACGCGCGAATACGTCGAA 799
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 706 TTCAGCAGATCCCGCTCGCGGCAACGCGTCG-----CAGTCGACCTTCGTCGAG 756
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 800 TGGCACTCAGAGAAATCTTTTCCACTCTATCTACACGATGACAGAGAGAGCTC 859
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 757 TGGAGCTCGGCGAACAACCGTGT-----CGACCTCCAGAACAGCGCCGAAC 801
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 860 TTTACCCTGTGAGCAACCATCGCTGCACAGACTGGCGCTTACGGCGGGTATGAGAAG 919
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 802 ACGAGCGCGGGAATGAGCGCTTCAACCTCTTCCCGCGGTTGATGTGAGCGCGTGTG 861
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 920 AAGATTGGCGTCCGAGGTCCAGGAATCTACTGGCCACCAAGGAATCTAGTTCGACGCTG 979
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 862 CGCAACAGCTCGAATCCGAGATCGAGAGTCTGCACATCCACGACTCATCTCGGGCGAC 921
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 980 GACAGCTATATATGCGGAAGGTATATGATGCGCTCGGAGCGATCGGCAACCCACAG 1039
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 922 CGCTTCAAAATCAAAGCAGACGTGTTCGTTTACAGCCGGGCGGTCCACAACGCGCAG 981
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1040 ATTCTCTATACCTGGGCTP-----CTCTGGGTACAGTCAAGTCCAGCCCAATGACTCG 1093
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 982 CTTCTCTGTAACCTTGGCTTTGGACAGCTGGCGCGCGGACCCCGGAAACCCGCGCAG 1041
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1094 TTGATCCCAACCTTGGGAGGTATACATCAGGAGCGGATGGCATTTTCCAGATAGTC 1153
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1042 TTGCTCGCGCTCCCTGGGAAGCTATATCCAGCAGTCTGCTCTCTCCGAGACCGTG 1101
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1154 TTGAGGAGGAATTCGTACAGCGTG-----CGCGAGATCTCTTATGACTGCG----- 1203
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1102 ATGAGCACCGAGCTCATCGACAGCGTCAAGTCCGACATGATCATCAGGGGCAACCCCTGGC 1161
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1204 -----ATGG 1207
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 3

US-09-023-731-2

; Sequence 2, Application US/09023731

; Patent No. 6291648

; GENERAL INFORMATION:

; APPLICANT: Kawamura, Yukio; Morita,

; APPLICANT: Akihiro; Izumo, Koji.; Saka, Tomohide.

; TITLE OF INVENTION: ANTITUMOR PROTEIN AND

; TITLE OF INVENTION: GENE ENCODING SAME

; NUMBER OF SEQUENCES: 20

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: MORGAN & FINNEGAN, L.L.P.

; STREET: 345 PARK AVENUE

; CITY: NEW YORK

; STATE: NEW YORK

; COUNTRY: USA

; ZIP: 10154

; COMPUTER READABLE FORM:

; MEDIUM TYPE: FLOPPY DISK

; COMPUTER: IBM PC COMPATIBLE

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: MICROSOFT WORD 97

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/023,731

; FILING DATE: 13-FEB-1998

; PRIOR APPLICATION DATA:

```

; APPLICATION NUMBER: JP 29275/1997
; FILING DATE: 13-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: KENNETH H. SONNENFELD
; REGISTRATION NUMBER: 33,285
; REFERENCE/DOCKET NUMBER: 3479-4000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1701
; TYPE: nucleic acid
; TOPOLOGY: linear
; STRANDEDNESS: Single
; MOLECULE TYPE: CDNA to RNA
; US-09-023-731-2

Query Match          7.2%; Score 152.6; DB 3; Length 1701;
Best Local Similarity 53.5%; Pred. No. 2.1e-25;
Matches 394; Conservative 0; Mismatches 304; Indels 39; Gaps 2;

QY 1106 CTGGGAGGTACATCAGGAGCAGCGGATGGCATTTTGGCCAGATAGTCTTGAGGCGAGGAA 1165
    || || || || || || || || || || || || || || || || || || || || || ||
DB 964 CTTGTCGCTACCTCAGCGAAGACATGCTGATGCTTTTGTTCAGATGCTTCTCAAGAGGGGC 1023
    || || || || || || || || || || || || || || || || || || || || || ||
QY 1166 TTCGTCGACGAGCGTGGCGACCATCTTATGACTGCCATGCTGGAAGAACGCCGTGCT 1225
    || || || || || || || || || || || || || || || || || || || || || ||
DB 1024 ATAGTCGATGCATCGCTACTGACCTT-----CGTTCGTCGGAAGTTGAG 1071
    || || || || || || || || || || || || || || || || || || || || || ||
QY 1226 CAACATATTGCCAAGAACCCGACAGATGCTGCCATTCCTTCGGGATCCGGAACCC 1285
    || || || || || || || || || || || || || || || || || || || || || ||
DB 1072 GCGCACAAAGAACGACCCGATGACGTGCTGCCATTCCTATTCACGAGCCTGAACCT 1131
    || || || || || || || || || || || || || || || || || || || || || ||
QY 1286 CAGGTAACAAACCCCATTTACAGAAACACCCCTGGCACACCCAGATTCACCGCATGCT 1345
    || || || || || || || || || || || || || || || || || || || || || ||
DB 1132 CAAGTGATGATTCCTGATACAGTCGCGACTTCCCTTGGCATGTTTCAGGTGCATCGCATGCA 1191
    || || || || || || || || || || || || || || || || || || || || || ||
QY 1346 TTTTCGTACGCTGCGTCTGCTGAGTGAGTCTCTGCTGTCATCTGTCACCTCGCTCGCTGG 1405
    || || || || || || || || || || || || || || || || || || || || || ||
DB 1192 TTTCTATATGGTGATGTTGGACCCCAAGCCGCGCTGTTGTCGTCGATCTGAGGTTT 1251
    || || || || || || || || || || || || || || || || || || || || || ||
QY 1406 TTTGGCGCAACCGACCCCTGAAGCAACAAACCTTTTGGTTTTC----- 1447
    || || || || || || || || || || || || || || || || || || || || || ||
DB 1252 TTCGCAAAATCAGATATTTCGAAGAAAATCAGTGACTTTCGGTCCGAACCTTAAGCTA 1311
    || || || || || || || || || || || || || || || || || || || || || ||
QY 1448 -----CAGAAGCATGTTCAAGACGGGTACAGTATGCCCGACCGGACGTTTCAGATAT 1498
    || || || || || || || || || || || || || || || || || || || || || ||
DB 1312 CCGGAGTGGGAAGCGGGTGTACAGACACTTATGGAATGCCACGCGGACATTCATGTC 1371
    || || || || || || || || || || || || || || || || || || || || || ||
QY 1499 CGACCCAGCAGCTCGCTCAAACTGAGAGCAAGGAAAATGATGGCCGATATGTGCGAAGTG 1558
    || || || || || || || || || || || || || || || || || || || || || ||
DB 1372 AAGCGGACCAACGCCGATGGAGACCGTGACAGAGGATGATGAATGATGACCAACGTC 1431
    || || || || || || || || || || || || || || || || || || || || || ||
QY 1559 GCGAGCACTTGGGAGGTTATTTGCCAGCTGCCCGCCGAGTTTATGATCCAGGCTTT 1618
    || || || || || || || || || || || || || || || || || || || || || ||
DB 1432 GCGAATCTGCTGGTGGTACCTTCTGCTCTTACCTCAATTTATGSCACCTGGTCTC 1491
    || || || || || || || || || || || || || || || || || || || || || ||
QY 1619 GCATTTCTTTCGCGGGGACTACTCGCATTTGCTTCGACAGGCAACTACAGTGCCTGAT 1678
    || || || || || || || || || || || || || || || || || || || || || ||
DB 1492 GTACTGCATCTACCGGGAACACTCTCGATCGGAGAGATGATCAAACTCTCTGCTGAT 1551
    || || || || || || || || || || || || || || || || || || || || || ||
QY 1679 ACAAATCGCTCTGGGACTTTTGGCAATCTTTTATGTTGAGGCAATGGCACCACGAG 1738
    || || || || || || || || || || || || || || || || || || || || || ||
DB 1552 CGACATCAAAAGTTCAAACTTCAACAATCTGTGGTGGGCGGGAATGGGTGCATCCA 1611
    || || || || || || || || || || || || || || || || || || || || || ||
QY 1739 AGGGCTTCGGGAGAACCCGACACTTACGTGATGTGCCAGCTATCAAGAGCCGAGG 1798
    || || || || || || || || || || || || || || || || || || || || || ||
DB 1612 GATGCGACTGCTCGAACCAGCCGACTCGTACGAGCGTGGGTATGCGCTCAAGGGTCTGAG 1671
    || || || || || || || || || || || || || || || || || || || || || ||
QY 1799 AGCATCATCAATACACT 1815
    || || || || || || || || || || || || || || || || || || || || || ||
```

Db 1672 GCTGTAGTCAATTACCT 1688

RESULT 4

```

US-09-149-476-24
; Sequence 24, Application US/09149476
; Patent No. 6420526
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 186 Human Secreted proteins
; FILE REFERENCE: P2002P1
; CURRENT APPLICATION NUMBER: US/09/149,476
; CURRENT FILING DATE: 1998-09-08
; EARLIER APPLICATION NUMBER: PCT/US98/04493
; EARLIER FILING DATE: 1998-03-06
; EARLIER APPLICATION NUMBER: 60/040,162
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,333
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/038,621
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,626
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,334
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,336
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,163
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/047,600
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,615
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,597
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,502
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,633
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,583
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,617
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,618
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,503
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,592
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,581
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,584
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,500
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,587
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,492
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,598
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,613
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,582
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,596
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,612
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,632
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,601
; EARLIER FILING DATE: 1997-05-23
```


	Query Match	4.2%	Score 88;	DB 4;	Length 2246;
	Best Local Similarity	77.9%;	Pred. No. 9e-11;	Mismatches 0;	Gaps 0;
	Matches 106;	Conservative			
Qy	1971	TGTCATTACTGAACCTGTACTTTGTCTGNAATAGTATTATGCACATATCATTCATGTTTTAAA	2030		
Db	2108	TGTCATTATATAAAGTGTACTTTGCCAAAAAA	2167		
Qy	2031	AA	2090		
Db	2168	AA	2227		
Qy	2091	AAAAAAAAAAAAAAAAAAAAA	2106		
Db	2228	AAAAAAAAAAAAAAAAAAAAA	2243		

EARLIER FILING DATE: 1998-07-30
EARLIER APPLICATION NUMBER: 60/095,486

;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: DE 196 21 588.9
;; FILING DATE: 29-MAY-1996
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: DE 196 36 917.7
;; FILING DATE: 11-SEP-1996
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: PCT/EP97/02793
;; FILING DATE: 28-MAY-1997
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Haley, Jr., James F.
;; REGISTRATION NUMBER: 27,794
;; REFERENCE/DOCKET NUMBER: AGREVO-9
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (212) 596-9000
;; TELEFAX: (212) 596-9090
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 2239 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: cDNA to mRNA
;; HYPOTHETICAL: NO
;; ANTI-SENSE: NO
;; ORIGINAL SOURCE:
;; ORGANISM: Triticum aestivum L.
;; STRAIN: cv. Florida
;; HAPLOTYPE: ca. 21 d Caryopses
;; IMMEDIATE SOURCE:
;; LIBRARY: cDNA library in pBluescript sk (-)
;; CLONE: TaSSS
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 3..2017
;;
US-09-196-390-1

Query Match 4.1%; Score 87; DB 4; Length 2239;
Best Local Similarity 67.2%; Pred. No. 1.5e-10;
Matches 123; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 1924 TCTGTCCTCCCTACCATGTGTGATGCTGACGATAGGCGTTGAAAGATTTTGTGTATTACTGA 1983
DB 2054 TGTGCTGCTGCTTGTAGCTGACAAATATTGACCTGTTGGAGATTTTATCTTGTGCT 2113
QY 1984 ACCTGTGACTTGTCTGAATAGTTATGGCACTATGATTCATGTTTAAAAAATAAAAAA 2043
DB 2114 GTTTTATTATCAAAAGAGGGGTTTCTCCGATTTTCATTAAAAAATAAAAAA 2173
QY 2044 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 2103
DB 2174 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 2233
QY 2104 AAA 2106
DB 2234 AAA 2236

RESULT 13
5168051-9
; Patent No. 5168051
; APPLICANT: DERYNCK, RIK M.A.; GOEDDEL, DAVID V.
; TITLE OF INVENTION: NUCLEIC ACID ENCODING TGF-B ITS USES
; NUMBER OF SEQUENCES: 21
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/389,929
; FILING DATE: 04-AUG-1989
; SEQ ID NO: 9:
; LENGTH: 2671
5168051-9

Query Match 4.1%; Score 86.6; DB 6; Length 2671;
Best Local Similarity 83.8%; Pred. No. 2e-10;

Matches 98; Conservative 0; Mismatches 19; Indels 0; Gaps 0;
QY 1990 ACTTGTCTCGAATAGTTATGGCACATATGATTCATGTTTAAAAAATAAAAAA 2049
DB 2544 ATTGTTCTAAACAATAAAGGCTTATTCTTAGGTCTAAAAAATAAAAAAATAAAAAA 2603
QY 2050 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 2106
DB 2604 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 2660

RESULT 14
US-09-489-847-23
; Sequence 23, Application US/09489847
; Patent No. 6476195
; GENERAL INFORMATION:
; APPLICANT: Rosen et al
; TITLE OF INVENTION: 98 Human Secreted Proteins
; FILE REFERENCE: P2031PI
; CURRENT APPLICATION NUMBER: US/09/489,847
; CURRENT FILING DATE: 2000-01-24
; EARLIER APPLICATION NUMBER: PCT/US99/17130
; EARLIER FILING DATE: 1999-07-29
; EARLIER APPLICATION NUMBER: 60/094,657
; EARLIER FILING DATE: 1998-07-30
; EARLIER APPLICATION NUMBER: 60/095,486
; EARLIER FILING DATE: 1998-08-05
; EARLIER APPLICATION NUMBER: 60/096,319
; EARLIER APPLICATION NUMBER: 60/095,454
; EARLIER FILING DATE: 1998-08-06
; EARLIER APPLICATION NUMBER: 60/095,455
; EARLIER FILING DATE: 1998-08-06
; NUMBER OF SEQ ID NOS: 376
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 23
; LENGTH: 1052
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-489-847-23

Query Match 4.1%; Score 85.6; DB 4; Length 1052;
Best Local Similarity 95.7%; Pred. No. 2.5e-10;
Matches 88; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 2015 ATGATTTCATGTTTAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 2074
DB 954 ATTTTAAATTTTAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 1013
QY 2075 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 2106
DB 1014 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 1045

RESULT 15
US-08-545-196B-10
; Sequence 10, Application US/08545196B
; Patent No. 6080577
; GENERAL INFORMATION:
; APPLICANT: MELKI, JUDITH
; APPLICANT: MUNNICH, ARNOLD
; TITLE OF INVENTION: SURVIVAL MOTOR NEURON (SMN) GENE: A GENE
; NUMBER OF SEQUENCES: 65
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH, LLP
; STREET: PO BOX 747
; CITY: FALLS CHURCH
; STATE: VA
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/545,196B
; FILING DATE: 19-OCT-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: FARACI, C. J.
; REGISTRATION NUMBER: 32,350
; REFERENCE/DOCKET NUMBER: 2121-110P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1582 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-545-196B-10

Query Match 4.0%; Score 85; DB 3; Length 1582;
Best Local Similarity 94.6%; Pred. No. 3.8e-10;
Matches 88; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Qy 2014 TATGATTCATGTTTAAA 2073
Db 1477 TTTAATTTTTTTTTAAA 1536
Qy 2074 AA 2106
Db 1537 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1569

Search completed: September 24, 2003, 05:33:21
Job time : 143 secs

Result No.	Score	Query 8		Length	DB	ID	Description
		Match	Score				
1	152.6	7.2	1701	10	US-09-911-969-2	Sequence 2, Appli	
2	94.4	4.5	4640	12	US-10-199-672-75	Sequence 75, Appl	
3	94.4	4.5	4640	12	US-10-187-749-75	Sequence 75, Appl	
4	94.4	4.5	4640	12	US-10-194-457-75	Sequence 75, Appl	
5	94.4	4.5	4640	12	US-10-184-642-75	Sequence 75, Appl	
6	94.4	4.5	4640	12	US-10-196-747-75	Sequence 75, Appl	
7	94.4	4.5	4640	12	US-10-173-689-75	Sequence 75, Appl	
8	94.4	4.5	4640	12	US-10-173-690-75	Sequence 75, Appl	
9	94.4	4.5	4640	12	US-10-173-691-75	Sequence 75, Appl	
10	94.4	4.5	4640	12	US-10-173-692-75	Sequence 75, Appl	
11	94.4	4.5	4640	12	US-10-173-694-75	Sequence 75, Appl	
12	94.4	4.5	4640	12	US-10-173-698-75	Sequence 75, Appl	
13	94.4	4.5	4640	12	US-10-173-699-75	Sequence 75, Appl	
14	94.4	4.5	4640	12	US-10-173-707-75	Sequence 75, Appl	
15	94.4	4.5	4640	12	US-10-174-569-75	Sequence 75, Appl	
16	94.4	4.5	4640	12	US-10-174-583-75	Sequence 75, Appl	

```

;
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 1701
;   TYPE: nucleic acid
;   TOPOLOGY: linear
;   (D) STRANDNESS: Single
; MOLECULE TYPE: cDNA to RNA
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-911-969-2

Query Match          7.2%;   Score 152.6;   DB 10;   Length 1701;
Best Local Similarity 53.5%;   Pred. No. 1.8e-21;
Matches 394;   Conservative 0;   Mismatches 304;   Indels 39;   Gaps 2;

QY 1106 CTGGGAGGTACATCAGCGAGCAGCGGATGGCATTTTGCACAGATAGTCTTTGAGCGAGGAA 1165
    || || || || || || || || || || || || || || || || || || || || || ||
Db 964 CTTGGTCGTACCTACCGCAACAGTCCATGACTTTTGTGCAGATCGTTCTCAGAGGGGC 1023

QY 1166 TTTCGTGCAGAGCGTGGCGACGATCCCTTATGAGCTGCCATGGTGGAAAGAACGGTTGCT 1225
    || || || || || || || || || || || || || || || || || || || || || ||
Db 1024 ATAGTCGATGCCATCGTACTGACCCCT-----CGCTTCGCTGCGAAGGTTGAG 1071

QY 1226 CAACATATTGCCAAGAACCCGACAGATGCACTGCCATTCCTGTTCCGGATCCGGAAACC 1285
    || || || || || || || || || || || || || || || || || || || || || ||
Db 1072 GCGCAACAAGAAAGAACCCCGGATGACGTCTGCCATTCCTATCCACGAGCGTGAACCT 1131

QY 1286 CAGGTAAACACCCATTTACAGAAGAACACCCCTGGCACACGACAGATTACACGCGATGCT 1345
    || || || || || || || || || || || || || || || || || || || || || ||
Db 1132 CAAGTGATGATCCGTACAGCTCGGACTTCCTTGGCATGTTCCAGTGCATCGCGATGCA 1191

QY 1346 TTTTCGTACGGTGCCGTCTCGGTCTGAGGTGGACTCTCGTGTCACTCGTCGACCTCGCGTGG 1405
    || || || || || || || || || || || || || || || || || || || || || ||
Db 1192 TTTCATATGGTGATGTTGGACCCAAAGCCGACCCCGGTGTTCTCGTGATCTGAGGTTT 1251

QY 1406 TTTGGCGCAACCGACCCCTGAAGCAACAACCTTTTGGTTTTTC----- 1447
    || || || || || || || || || || || || || || || || || || || || || ||
Db 1252 TTTCGGCAAAATCAGATATTGTGGAAGAAAAATCGAGTGACTTTTCGGTCCGAACCCCTAAGCTA 1311

QY 1448 -----CAGAACATGTTTCAAGAGCGGTACAGTATGCCGACGCGAGCTTCAGATAT 1498
    || || || || || || || || || || || || || || || || || || || || || ||
Db 1312 CGGAGTGGGAACGGGGTGTACAGACACTTATGGAAATGCCACAGCGGAAATTTCCATGTCT 1371

QY 1499 CGACCCAGCACTCGCTCAACAGTGAGAGCAAGAAATGATGGCCGATATGTGCGAAGTG 1558
    || || || || || || || || || || || || || || || || || || || || || ||
Db 1372 AAGCGACCAACGCCGATGGAGACCGTGCACAGAGATGATGATGATATGACCAACGCTC 1431

QY 1559 GCGAGCAACTTGGGAGGTATTATTTGCCACGTCCTCCCGCCAGTTTATGGATCCAGGCCTT 1618
    || || || || || || || || || || || || || || || || || || || || || ||
Db 1432 GCGAACATGCTGGTGGGTACCTTCTTGGCTCTACCTTCAATTTATGGCACCTGGTCTC 1491

QY 1619 GCATTCATCTTCGGGGGACTACTCGATGGCTTCGACNAGGCACTACAGTGGCTGAT 1678
    || || || || || || || || || || || || || || || || || || || || || ||
Db 1492 GTACTGCATCATCAGGGGAACACTACTCGGATCGGGACAGATGATCAAACTTCTGTTGCTGAT 1551

QY 1679 ACAACTCGCTGCTGGGACTTTGGCAATCTTTATGTTTCAGGCAATGCCACCATCAGG 1738
    || || || || || || || || || || || || || || || || || || || || || ||
Db 1552 CCGACATCAAGGTTTCATTAACATACTGTGGGTGCGCGGAATAGGGTGCAATTCCA 1611

QY 1739 ACGGGCTTCGGCGAGAACCCGACACTTAAGTCTGATGTGCCACGCTATCAAGAGCGCGAGG 1798
    || || || || || || || || || || || || || || || || || || || || || ||
Db 1612 GATGCGACTGCCGTCAACCCGACTCTGACGAGCGTCGCGGTATCGGCTCAAGGTCCTGAG 1671

QY 1799 AGCATCATCAATACACT 1815
    || || || || ||
Db 1672 GCTGTAGTCAATTAACCT 1688
    || || || || ||

```

RESULT 2

US-10-199-672-75
; Sequence 75, Application US/10199672
; Publication No. US2003018442A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian

RESULT 3

```

RESULT : 18
US-101-187-749-75
; Sequence 75, Application US/10187749
; Publication No. US2003015036A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.

```

```
; APPLICANT: Gurney,Austin L.
; APPLICANT: Pan,James
; APPLICANT: Smith,Victoria
; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William I.
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430RIC1
; CURRENT APPLICATION NUMBER: US/10/187,749
; CURRENT FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: US/10/052,586
; PRIOR FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059266
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/063120
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063121
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063486
; PRIOR FILING DATE: 1997-10-21
; PRIOR APPLICATION NUMBER: 60/063540
; PRIOR FILING DATE: 1997-10-21
; PRIOR APPLICATION NUMBER: 60/063540
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063541
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063541
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063544
; PRIOR FILING DATE: 1997-10-28
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 75
; LENGTH: 4640
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-187-749-75

Query Match          4.5%; Score 94.4; DB 12; Length 4640;
Best Local Similarity 72.9%; Pred. No. 2e-09;
Matches 137; Conservative 0; Mismatches 46; Indels 5; Gaps 1;

QY 1924 TCTGGTCCCTACCATTGTTGATGTCAGTATAGCGGTTTGAAGATTTTGTGTTACTGCA 1983
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4423 TATGCTCCATTTTATTTATAGTGTAAAGTTGTTATTTCCCTAAAGTTTGTGTTGTCGA 4482
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1984 ACCTGTACTTTGTCCTGA-----ATAGTTATGCGCACTATGATTCATGTTTAAAAA 2038
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4483 CAGTATCTTTTAAATGAGTCTTAAATAAAGGCATATTTGTTCTATGTTTAAAAA 4542
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2039 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2098
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4543 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4602
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2099 AAAAAAAA 2106
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4603 AAAAAAAA 4610
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 4
US-10-194-457-75
; Sequence 75, Application US/10194457
; Publication No. US20030153037A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
```

```
; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William I.
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430RIC296
; CURRENT APPLICATION NUMBER: US/10/194,457
; CURRENT FILING DATE: 2002-07-11
; PRIOR APPLICATION NUMBER: 10/052586
; PRIOR FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059266
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/063120
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063121
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063486
; PRIOR FILING DATE: 1997-10-21
; PRIOR APPLICATION NUMBER: 60/063540
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063541
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063544
; PRIOR FILING DATE: 1997-10-28
; Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 75
; LENGTH: 4640
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-194-457-75

Query Match          4.5%; Score 94.4; DB 12; Length 4640;
Best Local Similarity 72.9%; Pred. No. 2e-09;
Matches 137; Conservative 0; Mismatches 46; Indels 5; Gaps 1;

QY 1924 TCTGGTCCCTACCATTGTTGATGTCAGTATAGCGGTTTGAAGATTTTGTGTTACTGCA 1983
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4423 TATGCTCCATTTTATTTATAGTGTAAAGTTGTTATTTCCCTAAAGTTTGTGTTGTCGA 4482
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1984 ACCTGTACTTTGTCCTGA-----ATAGTTATGCGCACTATGATTCATGTTTAAAAA 2038
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4483 CAGTATCTTTTAAATGAGTCTTAAATAAAGGCATATTTGTTCTATGTTTAAAAA 4542
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2039 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2098
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4543 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4602
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2099 AAAAAAAA 2106
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4603 AAAAAAAA 4610
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 5
US-10-184-642-75
; Sequence 75, Application US/10184642
; Publication No. US20030157635A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
```


:	APPLICANT:	Goddard,Audrey			
:	APPLICANT:	Godowski,Paul J.			
:	APPLICANT:	Gurney,Austin L.			
:	APPLICANT:	Pan,James			
:	APPLICANT:	Smith,Victoria			
:	APPLICANT:	Watanabe,Colin K.			
:	APPLICANT:	Wood,William I.			
:	APPLICANT:	Zhang,Zemin			
:	TITLE OF INVENTION:	SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC			
:	FILE REFERENCE:	ACIDS ENCODING THE SAME			
:	CURRENT APPLICATION NUMBER:	US/10/173,694			
:	Prior Application removed - See File Wrapper or Palm				
:	NUMBER OF SEQ ID NOS:	612			
:	SEQ ID NO 75				
:	LENGTH:	4640			
:	TYPE:	DNA			
:	ORGANISM:	Homo Sapien			
:	US-10-173-694-75				
Query Match 4.5%: Score 94.4; DB 12; Length 4640;					
Best Local Similarity 72.9%; Pred. No. 2e-09;					
Matches 137; Conservative 0; Mismatches 46; Indels 5; Gaps 1;					
QY	1924	TCGTGGTCCCTACCATGTTGATGTGACGATAGCGCTTGAAAGATTTTTGTGTATTACTGA	1983		
Db	4423	TATGGCTCCATTTTATTTTAGTGTAAAGTGTGATTTCTCAAAGTTTGTGTTTGTGCGA	4482		
QY	1984	ACCTGTACTTTTGCTGA-----ATAGTTATGGCACATATGATTCATGTTTAAAAA	2038		
Db	4483	CAGTATCTTTTAATGTAGTCTTAAAAATAAAGGCATATGTTTCATGTTTAAAAA	4542		
QY	2039	AA	2098		
Db	4543	AA	4602		
QY	2099	AAAAAAAAA 2106			
Db	4603	AAAAAAAAA 4610			
RESULT 12					
US-10-173-698-75					
:	Sequence 75, Application US/10173698				
:	Publication No. US20030166108A1				
:	GENERAL INFORMATION:				
:	APPLICANT:	Baker,Kevin P.			
:	APPLICANT:	Chen,Jian			
:	APPLICANT:	Desnoyers,Luc			
:	APPLICANT:	Goddard,Audrey			
:	APPLICANT:	Godowski,Paul J.			
:	APPLICANT:	Gurney,Austin L.			
:	APPLICANT:	Pan,James			
:	APPLICANT:	Smith,Victoria			
:	APPLICANT:	Watanabe,Colin K.			
:	APPLICANT:	Wood,William I.			
:	APPLICANT:	Zhang,Zemin			
:	TITLE OF INVENTION:	SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC			
:	FILE REFERENCE:	ACIDS ENCODING THE SAME			
:	CURRENT APPLICATION NUMBER:	US/10/173,698			
:	Prior Application removed - See File Wrapper or Palm				
:	NUMBER OF SEQ ID NOS:	612			
:	SEQ ID NO 75				
:	LENGTH:	4640			
:	TYPE:	DNA			
:	ORGANISM:	Homo Sapien			
:	US-10-173-698-75				
Query Match 4.5%: Score 94.4; DB 12; Length 4640;					
Best Local Similarity 72.9%; Pred. No. 2e-09;					

```

; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C17
; CURRENT APPLICATION NUMBER: US/10/173,707
; PRIOR FILING DATE: 2002-06-17
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 75
; LENGTH: 4640
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-173-707-75

Query Match      4.5%; Score 94.4; DB 12; Length 4640;
Best Local Similarity 72.9%; Pred. No. 2e-09;
Matches 137; Conservative 0; Mismatches 46; Indels 5; Gaps 1;

QY 1924 TCTGGTCCCTACCATGTTGATGTGTACGATAGGCGTTGAAAGATTTTGTGTATTACTGA 1983
Db 4423 TATGGCTCCATTTTATTTTATAGTGTAAAGTTGTATTCTCTAAAGTTTGTGTTTGTGCGA 4482

QY 1984 ACCTGTACTTTTGTCTGA-----ATAGTTTATGGCAGCTATGATTCATGTTTAAAAA 2038
Db 4483 CAGTATCTTTTAAATGAGTCTTAAAAATAAAGGCATATTGTTTCATGTTTAAAAA 4542

QY 2039 AAAAAA 2106
Db 4543 AAAAAA 4610

QY 2099 AAAAAA 2106
Db 4603 AAAAAA 4610

Search completed: September 24, 2003, 05:56:49
Job time : 1401 secs

; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C39
; CURRENT APPLICATION NUMBER: US/10/174,569
; CURRENT FILING DATE: 2002-06-18
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 75
; LENGTH: 4640
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-174-569-75

Query Match      4.5%; Score 94.4; DB 12; Length 4640;
Best Local Similarity 72.9%; Pred. No. 2e-09;
Matches 137; Conservative 0; Mismatches 46; Indels 5; Gaps 1;

QY 1924 TCTGGTCCCTACCATGTTGATGTGTACGATAGGCGTTGAAAGATTTTGTGTATTACTGA 1983
Db 4423 TATGGCTCCATTTTATTTTATAGTGTAAAGTTGTATTCTCTAAAGTTTGTGTTTGTGCGA 4482

QY 1984 ACCTGTACTTTTGTCTGA-----ATAGTTTATGGCAGCTATGATTCATGTTTAAAAA 2038
Db 4483 CAGTATCTTTTAAATGAGTCTTAAAAATAAAGGCATATTGTTTCATGTTTAAAAA 4542

QY 2039 AAAAAA 2106
Db 4543 AAAAAA 4610

QY 2099 AAAAAA 2106
Db 4603 AAAAAA 4610

RESULT 15
US-10-174-569-75
; Sequence 75, Application US/10174569
; Publication No. US20030166111A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C39
; CURRENT APPLICATION NUMBER: US/10/174,569
; CURRENT FILING DATE: 2002-06-18
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 75
; LENGTH: 4640
; TYPE: DNA
; ORGANISM: Homo Sapien
```

```

US-10-174-569-75

Query Match      4.5%; Score 94.4; DB 12; Length 4640;
Best Local Similarity 72.9%; Pred. No. 2e-09;
Matches 137; Conservative 0; Mismatches 46; Indels 5; Gaps 1;

QY 1924 TCTGGTCCCTACCATGTTGATGTGTACGATAGGCGTTGAAAGATTTTGTGTATTACTGA 1983
Db 4423 TATGGCTCCATTTTATTTTATAGTGTAAAGTTGTATTCTCTAAAGTTTGTGTTTGTGCGA 4482

QY 1984 ACCTGTACTTTTGTCTGA-----ATAGTTTATGGCAGCTATGATTCATGTTTAAAAA 2038
Db 4483 CAGTATCTTTTAAATGAGTCTTAAAAATAAAGGCATATTGTTTCATGTTTAAAAA 4542

QY 2039 AAAAAA 2106
Db 4543 AAAAAA 4610

QY 2099 AAAAAA 2106
Db 4603 AAAAAA 4610

Search completed: September 24, 2003, 05:56:49
Job time : 1401 secs
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: September 24, 2003, 06:06:25 ; Search time 418 Seconds
(without alignments)
3680.635 Million cell updates/sec

Title: US-09-856-327-2

Perfect score: 3284

Sequence: 1 MSLSTEQMLRDPYRSMQING.....INTLKGGTDGKNTGHRNL 618

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 1678620 seqs, 1244745471 residues

Total number of hits satisfying chosen parameters: 3357240

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters: -DEV=xlh

-MODEL=frame.p2n.model
-O=/cgn2_1/USPTO_spool/US09856327/runat_23092003_153421_15588/app_query.fasta_1.775
-DB=PublishedApplications_NA -QFM=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-NAXLEN=2000000000 -USER=US09856327 @CGN_1.1.262 @runat_23092003_153421_15588
-NCPU=6 -ICPU=3 -NO_WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PublishedApplications_NA:

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
- 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
- 13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
- 14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
- 15: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
- 16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
- 17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	1091	33.2	1701	10	US-09-911-969-2 Sequence 2, Appli

2	177	5.4	1635	14	US-10-156-761-1918
3	177	5.4	9025608	14	US-10-156-761-1
4	148.5	4.5	1641	14	US-10-253-971-4
5	113	3.4	63158	12	US-10-292-198-1
6	112	3.4	1911	10	US-09-873-880-23
7	108.5	3.3	2346	14	US-10-353-445-8
8	108.5	3.3	3026	12	US-10-419-969-6
9	108	3.3	3431	14	US-10-128-714-75
10	108	3.3	3926	14	US-10-128-714-5075
11	107.5	3.3	9025608	14	US-10-156-761-1
12	107	3.3	4725	12	US-09-940-316B-24
13	106.5	3.2	3309400	10	US-09-738-626-1
14	106	3.2	3306	14	US-10-156-761-6845
15	106	3.2	68750	13	US-10-014-717-1
16	105.5	3.2	1542	14	US-10-156-761-1610
17	105	3.2	86114	15	US-10-080-170-648
18	104.5	3.2	1803	14	US-10-156-761-5001
19	104.5	3.2	32480	10	US-09-847-101B-23
20	104.5	3.2	32480	12	US-09-482-682-27
21	104.5	3.2	32798	12	US-10-424-638-1
22	104.5	3.2	34427	11	US-09-111-911-5
23	104.5	3.2	34555	12	US-10-117-982-479
24	104.5	3.2	35408	14	US-10-155-649-3
25	104.5	3.2	35871	9	US-09-956-335-2
26	104.5	3.2	35935	9	US-09-725-720-43
27	104.5	3.2	35935	10	US-09-782-378A-4
28	104.5	3.2	35935	10	US-09-782-378A-5
29	104.5	3.2	35935	12	US-09-739-007-43
30	104.5	3.2	35978	9	US-09-956-335-1
31	104.5	3.2	36620	11	US-09-952-060-30
32	104.5	3.2	37474	11	US-09-952-060-25
33	104.5	3.2	38519	11	US-09-952-060-28
34	104	3.2	4080	14	US-10-205-368-1
35	104	3.2	60196	14	US-10-205-032-1
36	103.5	3.2	1026	14	US-10-156-761-6467
37	103.5	3.2	36321	14	US-10-187-219A-1
38	103	3.1	5061	13	US-10-092-219-1
39	102.5	3.1	4611	10	US-09-801-368-321
40	102.5	3.1	4935	14	US-10-156-761-5059
41	102	3.1	3735	9	US-09-815-915-9
42	102	3.1	3735	12	US-10-393-316-9
43	102	3.1	5499	9	US-09-815-915-7
44	102	3.1	5499	12	US-10-393-316-7
45	102	3.1	26173	14	US-10-114-170-69

ALIGNMENTS

RESULT 1

US-09-911-969-2
; Sequence 2, Application US/09911969
; Patent No. US20020137896A1
; GENERAL INFORMATION:

APPLICANT: Kawamura, Yukio; Morita, Akihiko; Izumo, Koji.; Saka, Tomohide.
TITLE OF INVENTION: ANTITUMOR PROTEIN AND GENE ENCODING SAME

NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:

ADDRESSEE: MORGAN & FINNegan, L.L.P.
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154

COMPUTER READABLE FORM:

MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: MICROSOFT WORD 97

CURRENT APPLICATION NUMBER: US/09/911,969

FILING DATE: 24-Jul-2001

; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 09/023,731
 ; FILING DATE: 13-FEB-1998
 ; APPLICATION NUMBER: JP 29275/1997
 ; FILING DATE: 13-FEB-1997
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: KENNETH H. SONNENFELD
 ; REGISTRATION NUMBER: 33,285
 ; REFERENCE/DOCKET NUMBER: 3479-4000
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 758-4800
 ; TELEFAX: (212) 751-6849
 ;
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1701
 ; TYPE: nucleic acid
 ; TOPOLOGY: linear
 ;
 ; MOLECULE TYPE: CDNA TO RNA
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
 US-09-911-969-2

Alignment Scores:
 Pred. No.: 2,05e-110 Length: 1701
 Score: 1091.00 Matches: 248
 Percent Similarity: 57.04% Conservative: 84
 Best Local Similarity: 42.61% Mismatches: 184
 Query Match: 33.22% Indels: 66
 DB: 10 Gaps: 15

US-09-856-327-2 (1-618) x US-09-911-969-2 (1-1701)

```

; 37 AspValPheIleAlaGlySerGlyProIleGlyAlaThrTyrAlaLysLeuCysValGlu 56
; ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
; 97 GATGTTTTCATTCTGGCAGTGGTCCCATAGTGTACTGTACCGCCGCACATCATGTAC 156
;
; 57 -----AlaGlyLeuArgValValMetValGluIleGlyAlaAlaAspSerPheTyrAla 74
; :::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
; 157 AATACCTCAACTACAAAGGTTTACATGGCGGAATAGGTCTCAAGATAAC----- 207
;
; 75 ValAsnAlaGluGluGlyThrAlaValProTyrValProGlyTyrHisLysLysAsnGlu 94
; ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
; 208 -----CCTGTCATC-----GGGCCCATCACAGAATCTC 237
;
; 95 IleGluPheGlnLysAspIleAspArgPheValAsnValIleLysGlyAlaLeuGlnGln 114
; ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
; 238 AATAAGTTTCAGAAACACATTCACAAGTTTGTGAATATCATCAACGGTGCCCTCCAGCG 297
;
; 115 ValSerValProValArgAsnGlnAsnValProThrLeuAspProGlyAlaTrpSerAla 134
; :::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
; 298 ATTTTCGATTTCGCCATCGGACACCTACCAGCCCATCTCGCTGTAGCAGCGTGG---GGG 354
;
; 135 ProGlySerSerAla-----IleSerAsnGlyLysAsnProHisGlnArg 150
; ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
; 355 CCGCCCATCATCTCCTCCGAAGCCAGCTCGTGATTATGGACACAAATCCGAATCAGAG 414
;
; 151 GluPheGluAsnLeuSerAlaGluAlaValThrArgGlyValGlyMetSerThrHis 170
; ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
; 415 GCGGGCTGAACCTTCCCGTAGCGTGTCACTAGGACAGATCGGGGGGATGGGACCCAC 474
;
; 171 TrpThrCysSerThrProArgIleHisProPrometGluSerLeuProGlyIleGlyArg 190
; ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
; 475 TGGACTTGGCGGTGCTCTACTCCACATGAC----- 504
;
; 191 ProLysLeuSerAsnAspProAlaGluAspAspLysGluTrpAsnGluLeuTyrSerGlu 210
; :::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
; 505 GAAGAGAGGGTCAACACCCAGT---GACACAGCAGGAGTTCGACCACTGCTCGAAGCT 561
;
; 211 AlaGluArgLeuIleGlyThrSerThrLysGluPheAspGluSerIleArgHisThrLeu 230
; ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
; 562 GCTAATACATTCGCTCAACGTTTCACAGCGCAGTACGACGATTCATCCGTCAGATAGTT 621
  
```

Db 1624 TCACACCGGACTCGTACGAGCGTCGCTATCGCGCTCAAGGGTGTCTAGTCAAT 1683
 Qy 602 ThrLeu 603
 Db 1684 TACCTT 1689

RESULT 2

US-10-156-761-1918
 ; Sequence 1918, Application US/10156761
 ; Publication No. US20030119018A1
 ; GENERAL INFORMATION:
 ; APPLICANT: OMURA, SATOSHI
 ; APPLICANT: IKEDA, HARUO
 ; APPLICANT: ISHIKAWA, JUN
 ; APPLICANT: HORIKAWA, HIROSHI
 ; APPLICANT: SHIBA, TADAYOSHI
 ; APPLICANT: SAKAKI, YOSHIYUKI
 ; APPLICANT: HATTORI, MASAHIRA
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 ; FILE REFERENCE: 249-262
 ; CURRENT APPLICATION NUMBER: US/10/156,761
 ; PRIOR FILING DATE: 2002-05-29
 ; PRIOR APPLICATION NUMBER: JP 2001-204089
 ; PRIOR FILING DATE: 2001-05-30
 ; PRIOR APPLICATION NUMBER: JP 2001-272697
 ; PRIOR FILING DATE: 2001-08-02
 ; NUMBER OF SEQ ID NOS: 15109
 ; SEQ ID NO 1918
 ; LENGTH: 1635
 ; TYPE: DNA
 ; ORGANISM: Streptomyces avermitilis
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1)..(1635)
 US-10-156-761-1918

Alignment Scores:

Pred. No.: 2,546-09 Length: 1635
 Score: 177.00 Matches: 144
 Percent Similarity: 32.60% Conservative: 63
 Best Local Similarity: 22.68% Mismatches: 249
 Query Match: 5.39% Indels: 179
 DB: 14 Gaps: 26

US-09-856-327-2 (1-618) x US-10-156-761-1918 (1-1635)

Qy 37 AspValPheIleAlaGlySerGlyProIleGlyAlaThrTyrAlaLysLeuCysValGlu 56
 Db 73 GAGGTCATCATATTGGCACCAGCGGTGCGGCGGTGGCAGCGCTCGCCACCGACTGGCCCCC 132
 Qy 57 AlaGlyLeuArgValValMetValGluIleGly-----AlaAlaAspSer 71
 Db 133 ACCGGGAACGGATCCTCTGCTCGAACCGGTGTACTACCTGCCCGCGGAACGCCACAAAC 192
 Qy 72 PheTyrAlaValAsnAlaGluGlyThrAlaValProTyrValProGlyTyrHisLys 91
 Db 193 TGG-----GAATCCACCGCGGTGCTTCTCAAGGGCAAAATACCGC 231
 Qy 92 LysAsnGluIleGluPheGlnLysAspIleAspArgphe-----ValAsnValIle 108
 Db 232 GCTCGGAGTTCTTGGTACGACAAACACCGGAACACAGTTCGCCCGCCGAAGTCAATTACTAC 291
 Qy 109 LysGly-AlaLeuGlnValSerValProValArgAsnGlnAsnValProThrLeuAs 128
 Db 292 GTCGGCGGAACACCAAGTTCTACCGCGCGCGGTCTTCCTCGCTCGCGCCCGAGGACTTC 351
 Qy 128 pProGlyAla-----TrpSerAlaProGlySerSerAlaIleSerAsnG1 144
 Db 352 GCGGAACCTCGGCCACACGAGCGCATCTCCCGCGCTGCGCGCTGAGCTACGAGGAATC 411
 Qy 144 yLysAsnProHisGlnArgGluPheGlnAsnSerAlaGluAlaValThrArgGlyVa 164

Db 412 GAGCGGTACTACACGAGCGCCGAACACCTCTACCTCG----- 448
 Qy 164 lGlyGlyMetSerThrHisTrpThrCysSerThr---ProArgIleHisProProMetG1 183
 Db 449 -----TCCACGGCGCGCACCGCGAAGACCCGACCGAGGGTC---CCACGAGCGC 495
 Qy 183 uSerLeuProGlyIleGlyArgProLysLeuSerAsnAspProAlaGluAspLysG1 203
 Db 496 CAGTACGGCT---ACCCGCGGTCCAGCAGCAGCGCGCATCCAGCACTACGACCCACGAC 552
 Qy 203 uTrpAsnGluLeuTyrSerGluAlaGluArgLeuIleGlyThrSerThrLysGluPhe-A 223
 Db 553 CTGG-----AGAAAGCGGGCTCGACCCCTTCACCTTCACCTTCG 588
 Qy 223 spLusSerIleArgHisThrLeuValLeuArgSerLeuGlnAspAlaTyrLysAspArg 243
 Db 589 ATCGGGGTCAACCTCACCCA-----GGACGACCGGGCGGGCCACCCACGCGC 636
 Qy 243 lArgIlePhe-----ArgProLeuProLeu----- 251
 Db 637 AGCGCTGCATCCGCTGCGACCGCGTCGACGGCTTCCCTGTCTGTGTCGCGCGAAGTCC 696
 Qy 252 ----AlaCysHisArgLeuLysAsnAlaPro-GluTyrValGluTrpHisSerAlaGlu 269
 Db 697 GACGCGCAGGTCACTGTCTGCACCCCGCCCTGGAGCAGCGC-----AACGTGAG 747
 Qy 270 AsnLeuPheHisSerIleTyrAsnAspLysGlnLysLysLeuPheThrLeuLeuThr 289
 Db 748 ATGCTCACCATGG-----GATGTGCGGGCGCTCGACACGCGCGCACC 792
 Qy 290 AsnHisArgCysThrArgLeuAlaLeuThrGlyTyrGluLysLysIleGlyAlaAla 309
 Db 793 GCGGAGGTGTCACTCGCTCGCGACGGTGGG-----CACGGGCGCGCT 840
 Qy 310 GluValArgAsnLeuLeuAlaThrArgAsnProSerSerGlnLeuAspSerTyrIleMet 329
 Db 841 TCCACCGTGGAG-----TTCACG 858
 Qy 330 AlaLysValTyrValLeuAlaSerGlyAlaIleGlyAsnProGlnIleLeuTyrAsnSer 349
 Db 859 GCGGACATCGTGTGCTGCGCGCGCGCTCACTCGCGCTCTGTGTGCTGCGCTTCG 918
 Qy 350 GlyPheSerGlyLeuGlnValThrProArgAsnAspSerLeuIleProAsnLeuGlyArg 369
 Db 919 GCC----- 921
 Qy 370 TyrIleThrGluGlnProMetAlaPheCysGlnIleValLeuArgGlnGluPheValAsp 389
 Db 921 ----- 921
 Qy 390 SerValArgAspAsp-----ProTyrGlyLeuProTyrTrpLysGluAlaValAlaGln 407
 Db 922 -----GATGACCGGCATCCGCGCGCTGGCCACAGCTCGGGCGGTGGGCGCG 972
 Qy 408 HisIleAlaLysAsnProThrAspAlaLeuProIleProPheArgAspProGluProGln 427
 Db 973 CACTACATGCGGCACAAACCTGCGCTGTGCGCGTGTCCCAAGGAACCGAACACAC 1032
 Qy 428 ValThrThrProPheThrGluGluHisProTyrPHis----- 439
 Db 1033 AAATTCAGAGACCCCTGGCGCTGCAGCAGCTGTGTACCTGGGATCCGACGACTGGGACTAC 1092
 Qy 440 -----ThrGlnIleHisArgAspAla 446
 Db 1093 CCTCTCGGGGCATCCAGATGCTCGGCAAGTCCGACGCGCAGCATCCACGCGCGAGGG 1152
 Qy 447 PheSerTyr---GlyAlaValGlyProGluValAspSerArgValIle----- 461
 Db 1153 CCCCGCTGGCGCGCGCTCGCCCGCAGCATGCGCTTCGAGGTACTTCGCCCCACACGCG 1212
 Qy 462 ValAspLeuArgTyrPheGlyAlaThrAspProGluAlaAsnAsnLeuLeuValPheGln 481
 Db 1213 CTCGACTTCTGGCTGTGGCGAGAGACCTGCCCTCGCGGAGAACCGGTCTACCTCCCTGGAC 1272


```
Qy 390 SerValArgAspAsp-----ProTyrGlyLeuProThrTrpTrpLysGluAlaValAlaGln 407
Db 2359983 -----GATGACGGCATCCGCGAGCGCTGCCAACAGCTCGGGCGTGGTGGCGG 2360033
Qy 408 HisIleAlaLysAsnProThrAspAlaLeuProIleProPheArgAspProGluProGln 427
Db 2360034 CACTACATGCGCGCAACAACCTGGCCCTGATGGCCGTGTCCAGGACCGAACGACACC 2360093
Qy 428 ValThrThrProPheThrGluGluHisProThrHis----- 439
Db 2360094 AAATCCAGAGACCCCTGGCCCTGCAGCAGTGGTACCTGGGATCCGACGACTGGGACTAC 2360153
Qy 440 -----ThrGlnIleHisAspAla 446
Db 2360154 CCTCGCGGCATCCAGATCTCGGCAAGTCCGACGCGGAGATCCACGCGGAGGCG 2360213
Qy 447 PheSerTyr---GlyAlaValGlyProGluValAspSerArgValIle----- 461
Db 2360214 CCCCGCTGGCGGGCGGCGTCCGCCGACATCGCTGAGGTACTCGCCACACGCG 2360273
Qy 462 ValAspLeuArgTrpPheGlyAlaThrAspProGluAlaAsnAsnLeuValPheGln 481
Db 2360274 GTCGACITCTGGCTGTCGGGAGAGGACCTGCCCTCGCGAGAACCGCGTCACTCGGAC 2360333
Qy 482 AsnAspValGlnAspGlyTyrSerMetProGlnProThrPheArgTyrArgProSerThr 501
Db 2360334 GGGGACGCGGCATCCATCTGGCGCTGCACGAG-----AAGAACAACATC 2360378
Qy 502 AlaSerAsnValArgAlaArgLysMetMetAlaAspMetCysGluValAlaSerAsnLeu 521
Db 2360379 GCGGGGTGAAGCGCTGCGGCACAACTACAGGGCATG-----CTCAGCCACTTG 2360429
Qy 522 GlyGly-----TyrLeuProThrSerProGln 531
Db 2360430 GGCATGACGAGGACCATCTGCTGTCGACAGCATCTACCTGCACAGGGCATGCC--- 2360486
Qy 532 PheMetAspProGlyLeuAlaLeuHisLeuAlaGlyThrThrArgIleGlyPheAspLys 551
Db 2360487 -----ATCGCGGCACCGCGCATCAGCGGGCACGGTCCGCTTCGCGCGGACCC 2360537
Qy 552 AlaThrThrValAlaAlaAsnAsnSerLeuValTrpAspPheAlaAsnLeuTyrValAla 571
Db 2360538 GCCGACTCGCGCTGAGTCACTGCAAGCCCGACGCTCGACACCTCTAGTCGTCTC 2360597
Qy 572 GlyAsnGlyThrIleArgThrGlyPheGlyGluAsnProThrLeuThrSerMetCysHis 591
Db 2360598 GACACGAGCTTCTTCCCGAGCATCGCGCGGTGAATCCCTGCTGACCGCATCGCCAAC 2360657
Qy 592 AlaIleLysSerAlaArgSerIleLeuAsnThrLeuLys 604
Db 2360658 GCCCTGGGGTTCGGCGACCATCTCGCGGAGGACTGGGG 2360696
```

RESULT 4

```
US-10-253-971-4
; Sequence 4, Application US/10253971
; Publication No. US20030070192A1
; GENERAL INFORMATION:
; APPLICANT: GEORGES, FAWZY
; APPLICANT: DONG, JIN-ZHUO
; APPLICANT: KELLER, WILF
; APPLICANT: HUSSAIN, ATTA A. K.
; APPLICANT: SELVARAJ, GOPALAN
; APPLICANT: DATLA, RAJU
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR MODULATING SECONDARY PLANT
; FILE REFERENCE: 73776-159D
; CURRENT APPLICATION NUMBER: US/10/253,971
; CURRENT FILING DATE: 2002-09-25
; PRIOR APPLICATION NUMBER: US 60/072156
; PRIOR FILING DATE: 1998-01-22
; PRIOR APPLICATION NUMBER: US 09/012453
; PRIOR FILING DATE: 1998-01-23
; NUMBER OF SEQ ID NOS: 7
```

```
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1641
; TYPE: DNA
; ORGANISM: Arthrobacter pascens
US-10-253-971-4
Alignment Scores:
Pred. No.: 3.63e-06 Length: 1641
Score: 148.50 Matches: 128
Percent Similarity: 32.90% Conservative: 75
Best Local Similarity: 20.75% Mismatches: 238
Query Match: 4.52% Indels: 177
DB: 14 Gaps: 27
US-09-856-327-2 (1-618) x US-10-253-971-4 (1-1641)
Qy 37 AspValPheIleAlaGlySerGlyProIleGlyAlaThrTyr---AlaLysLeuCysVal 55
Db 43 GACTACATCATCATCGCGCGGTTCGCCGAGCGGCGAGTCCGCCGCTGAGCGAG 102
Qy 56 GluAlaGlyLeuArgValMetValGluIleGlyAlaAlaAspSerPheTyrAlaVal 75
Db 103 GAGCCACCGGTTCGTGGCGCTGTGGAGCGCGCGGACGAC----- 147
Qy 76 AsnAlaGluGlyThrAlaValProTyrValProGlyTyrHisLysLysAsnGluIle 95
Db 148 -----CGCGCGTTCGCGAGGTA----- 165
Qy 96 GluPheGlnLysAspIleAspArgPheValAsnValIleLysGlyAlaLeuGlnGlnVal 115
Db 166 -----CTCGAGCTCGACGCTGGATGGTGTGTAATCCGCTACGAC----- 210
Qy 116 SerValProValArgAsnGlnAsnValProThrLeuAspProGlyAlaTrpSerAlaPro 135
Db 211 -----TGGACTATCCCG 222
Qy 136 -----ProGlySer-SerAlaIleSerAsnGlyLysAsnProHisGlnArgGluPheG1 153
Db 223 ATCGAACCGCAGGAGAACGCACTCTTCATCGCCCGCCCGCGCGAGATCATGGGT 282
Qy 153 uAsnLeuSerAlaGluAlaValThrArgGlyValGlyGlyMetSerThrHisTrpThrCy 173
Db 283 GGCTGCTCCAGCAACAACCTCTGCTGCTTCCTGGCGCGCGCGAGAGACTGGACGAG 342
Qy 173 sSerThrProArgIleHisProMetGluSerLeuProGlyIleGlyArgProLysLe 193
Db 343 TGGGAGTCCAAAGTACGGCGCCCGCTGGGAACGCTG----- 379
Qy 193 uSerAsnAspProAlaGluAspAspLysGluTrpAsnGluLeuTyrSerGluAlaGluAr 213
Db 380 ---AGTCCGCTGCGCGCTGTACAGCGGCTGGAGACCAACGAGGACGCGCGCGGAC 435
Qy 213 gLeuIleGlyThrSerThrLys----- 220
Db 436 GCGCGCACCGCAGCGGACTCAGCGCGCGTGCACCTGATGAACGCTGCCCGCGCGACCCC 495
Qy 221 -----GluPheAspGluSerIleArgHisThrLeuValLeuArgSerLeuG1 236
Db 496 GCGCGGCTGCGACTCTCGAGCGCTCGAACAGGACGAGCATTCGCGCGCGGAGTTCAAC 555
Qy 236 nAspAla-TyrLysAspArgGlnArgIlePheArgPro-----LeuProLeuAlaCysH 254
Db 556 ACCGCGACCGCGTGAATCAATGCGCCCACTTTTCCAGATCACACCGCGCGGACGCG 615
Qy 254 is-ArgLeuLysAsnAlaProGluTyrValGluTrpHisSerAlaGluAsnLeuPheHis 273
Db 616 ACCGCTTCTCCAGCTCGGTCTCTACATC-----CAG 648
Qy 274 SerIleTyrAsnAspAspLysGlnLysLysLeuPheThrLeuLeuThrAsnHisArgCys 293
Db 649 CCGATCATC-----GAGCGCGGGAACCTTCACTCCCTGCTGACCGGGTTCGCGGCC 696
```


Db 59566 CGGCTGCGCGCTTCGAGGCG-----AAGCAG-----TCCTTGCT 59534
QY 288 uThrAsnHisArgCysThrArgLeuAlaLeuThrGlyGlyTyrGluLysLysIleGlyAl 308
Db 59533 GGTCAACTACCTGCGCGCGCATCGCC-----CATGGGCG 59498
QY 308 aAla-----GluValArgAsnLeuLeuAlaThrArgAsnProSerSerG 323
Db 59497 GCGCATCGCGCGCTGTCAGAGGTTCAGCACCTCTCGCGCACCCCGCGGCTCTACCG 59438
QY 323 nLeu-----AspSerTyr-----II 328
Db 59437 GGTGCATACAACTGTCAGCAGCAGCACTCCGCTCCAGCGCGGCGAGCGGTCTCT 59378
QY 328 eMetAlaLysValTyrValLeuAlaSerGlyAlaIleGlyAsnProGlnIleLeuTyrAs 348
Db 59377 CGAGCGAAGATCGTGGTGATGGCGCGCGCGCGCGCGCGCGGTGATCCTCCAGCG 59318
QY 348 nSerGlyPheSerGlyLeuGlnValThrProArgAsnAspSerLeuIleProAsnLeuG 368
Db 59317 CAGCGAG---GCCACCTGGGCGCACCATGCCAGCA-----GCGGTGCG 59279
QY 368 yArgTyrIleThrGluGlnProMetAlaPheCysGlnIleValLeuArgGlnGluPheVa 388
Db 59278 ACCTACTTCTCCGGCAACGCGGAGCGG-----CTCAACCGCGCATCAT 59234
QY 388 lAspSerValArg-----AspAspProTyrGlyLeuProTyrTrp 401
Db 59233 CGACGAGCCAGGCGCGGAGCTGTCGGCTGACCGCGGCGCGGCTCGCTAC-- 59176
QY 401 pLysGluAlaValAlaGlnHisIleAlaLysAsnProThrAspAla-LeuProIlePro 421
Db 59175 -----GGGGCAACCATGATCGCAAGGGGCCCGCTGGCGAGCTGGACAGGCT 59126
QY 421 heArgAspProGluProGlnValThrThrProPheThrGluLysHisProTrpHisThrG 441
Db 59125 GGACGGCTGTTGCCCG-----AGTACTCCCGCTACT-----CCCTGGAACAGCT 59081
QY 441 lInIleHisArgAspAlaPheSerTyr-----GlyAlaValGlyProG 455
Db 59080 CTACTTCCCGCGCGGCTCGGCACG-ATCCTCGCCCGAGTGCCCGCGCGCGCGCCCT 59022
QY 455 luValAspSerArgValIleValAspLeuArgTyrPheGlyAlaThrAspProGlu---- 473
Db 59021 CC-----TGGTTCGGCAAGGAGAGAGAGATCC 58992
QY 474 -----AlaAsnAsnLeuLeuValPheGln-----AsnAspValGlnAspGlyT 488
Db 58991 TGAACGAGTGACCTCTCGGCTCAGGATCTTCAGCATGATCGAGGACGACAAAGAGCG 58932
QY 488 yrSerMetProGlnPro----- 493
Db 58931 TCTTCGCGCCCTCCCGCCACCGGTACGCCACCGCATCTCCCGCAGATGCTCGGCC 58872
QY 494 -----ThrPheArgTyrArgProSer-----ThrAlaSerAsnV 505
Db 58871 GGGGCAACCTCGCTACGACCGAGCAACACCCCTGGCGCCCTGGCGCGCATCCGACG 58812
QY 505 alArgAlaArgLysMet-----MetAlaAspMetCysGluValAlaSerA 520
Db 58811 CCGAGGTCAAGAGATCTCTGGAGAGGACGCGCTGGCCAAAGGTATCGCTGGACCAACG 58752
QY 520 snLeuGlyGlyTyrThrSerProProGlnPheMetAspProGlyLeuAlaLeuH 540
Db 58751 ACCTCGTCGCG-----GCCCTACACCGTCC 58728
QY 540 lsLeuAlaGlyThrThrArgIleGlyPheAspLysAlaThrValAlaAspAsnAsns 560
Db 58727 ACCCGCTCTCTCTCGCGCGATGGCGAGACCCACACATCCGCCCTGGACGACACCA 58668
QY 560 erLeuValThrAspPheAlaAsnLeuTyrValAlaGlyAsnGlyThrIleArgThrGlyP 580
Db 58667 ACGAGCTCCGGGACCAACCGCGGATCTTCGTCAGCGAGGATCTCGGTTCCCGCGCGCG 58608

QY 580 heGlyGluAsnProThrLeuThrSerMetCysHisAlaIleLysSerAlaArgSerIleI 600
Db 58607 TGACGGTCAATCCCGCGATGACCATCGCGCGCTCGCCGAACGCGGATCGCGGGATCG 58548
QY 600 leAsnThrLeuLysGly-----GlyThrAspGlyL 610
Db 58547 TCGCGCGCGCGAGAGCGGGGTATCTCCGTCACTACGCGCAGCGCGCGCGCGGCT 58488
QY 610 ysAsnThrGlyGluHisArgAsnLeu 618
Db 58487 CCACGAGCGCGCGAGAGGTTCTT 58462

RESULT 6

US-09-873-880-23
; Sequence 23, Application US/09873880
; Patent No. US20020123118A1
; GENERAL INFORMATION:
; APPLICANT: Sewalt, Vincent
; APPLICANT: Falco, S. Carl
; APPLICANT: Allen, Stephen M.
; TITLE OF INVENTION: GLYCINE METABOLISM ENZYMES
; FILE REFERENCE: BB192 US CIP
; CURRENT APPLICATION NUMBER: US/09/873,880
; PRIOR FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: 09/363,321
; PRIOR FILING DATE: July 28, 1999
; PRIOR APPLICATION NUMBER: 60/094,839
; PRIOR FILING DATE: July 31, 1998
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 23
; LENGTH: 1911
; TYPE: DNA
; ORGANISM: Zea mays
US-09-873-880-23

Alignment Scores:

Pred. No.:	0.0498	Length:	1911
Score:	112.00	Matches:	132
Percent Similarity:	32.50%	Conservative:	88
Best Local Similarity:	19.50%	Mismatches:	214
Query Match:	3.41%	Indels:	244
DB:	10	Gaps:	33

US-09-856-327-2 (1-618) x US-09-873-880-23 (1-1911)

QY 2 SerLeuSerThrGluGlnMetLeuArgAspTyrProArgSerMetGlnIleAsnGlyGln 21
Db 84 TCGCTCAACTCTCCACATGACTACTGAGTAT----- 116
QY 22 IleProLysAsnAlaIleHisLysThrTyrGlyAsnAspGlyValAspValPheIleAla 41
Db 117 CTTCCTCGCTCTGCC-----AGTCGCGCTACGACTATATCATCGTA 158
QY 42 GlySerGlyProIleGlyAlaThrTyrAlaLysLeuCysValGluAlaGly----- 58
Db 159 GGTGTGCGCAGCGCTGGA-----TGTGTCTGGCTTCCCGCTATCC 200
QY 59 -----LeuArgValValMetValGluIleGlyAlaAspSerPheTyr 73
Db 201 TCCTACCTTCCTGAGCGCAAGGTTTATGATTGAGCTGGCCCTTCAGAC-----TTC 254
QY 74 AlaValAsnAlaGluGluGlyThrAlaValProTyrValProGlyTyrHisLysLysAsn 93
Db 255 GGTCTCAAC-----AAT 266
QY 94 GluIleGluPheGlnLysAspIleAspArgPheValAsnValIleLysGlyAlaLeuGln 113
Db 267 GTCCTGAACCTTCGCGAG-----TGGCTGTCTCTCTTGGTGGTGTCTC--- 311
QY 114 GlnValSerValProValArgAsnGlnAsnValProThrLeuAspProGlyAlaTrpSer 133

```
Db 312 -----GACTACGATTATCCACCACTGAG----- 335
Qy 134 AlaProGlySerSerAlaIleSerAsnGlyLysAsnProHisGlnArgGluPheGlu 153
Db 336 CAGCCCAATGGCAACGCCACATCCGA-----CACTACGT----- 371
Qy 154 AsnLeuSerAlaGluAlaValThrArgGlyValGlyGlyMetSerThrHisTrpThrCys 173
Db 372 -----GCAAAAGTCCTGGTGGATGCTCTCTCACAACACTCTC 410
Qy 174 SerThrProArgIleHisProPro-----MetGluSerLeuProGly 187
Db 411 ATCTCTTCGCTCC-TTCCGCCACGACATGCGATGCTGGTGCACCAAGCTGCAAGG 469
Qy 188 IleGly-----ArgProLys 192
Db 470 CTGGGACTCGAGACCGTTATGCGCAACGTTGCAACTTGGCGAACCACTGAACCTGT 529
Qy 193 LeuSerAsnAspPro-AlaGluAspAspLysGluTrpAsnGluLeuTyrSerGluAlaG1 212
Db 530 TCATCCCGCTCACCGTAACAGCTCACCAAGGACTGGTCAAGGCGCTGCTCCGAGGCCAT 589
Qy 212 u-----ArgLeuIleGlyThrSerThrLysGluPheAspGluSerIleArgHisThrLe 230
Db 590 GGCATATCCATCCATCCACGACTTCAATCACGAGATTTCAGAGGGAGCTGTGACCCA 649
Qy 230 uValLeuArgSerLeuGlnAspAlaTyrLysAspArgGlnArgIlePheArgProLeuPr 250
Db 650 GGTGTCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 682
Qy 250 oLeuAlaCysHisArgLeuLysAsnAlaProGluTyrValGluTrpHisSerAlaGluAs 270
Db 683 TGACACCGGCCACCGCAGCAGTCTTCCGCTGCCTATATC----- 722
Qy 270 nLeuPheHisSerIleTyrAsnAspLysGlnLysLysLeuPheThrLeuLeuThrAs 290
Db 723 -----CACCTATCTCTGCGGCGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 773
Qy 290 nHisArgCysThrArgLeuAlaLeuThrGlyGlyTyrGluLysLysIleGlyAlaAlaG1 310
Db 774 -----GAGGCCCATGTCTCAAGGTCTATC----- 796
Qy 310 uValArgAsnLeuLeuAlaThr-----ArgAsnProSerSerGlnLeuAspSe 326
Db 797 CGTCGAAATAGCTGTGCCACTGGCATCAATGTCACTCTCAAGTCAAGCGAGAGACAC 856
Qy 326 rTyrIleMetAlaLysValTyrValLeuAlaSerGlyAlaIleGlyAsnProGlnIleLe 346
Db 857 TCTGAACGCCCGCAGAGAGATCATCTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 916
Qy 346 uTyrAsnSerGly-----PheSerGlyLeuGlnValThrProArgAs 360
Db 917 CTCCCACTCTGTTATGGACCAAGGCCAGCTTGAAGACCTGAAGATTCTCTGTCTCAA 976
Qy 360 nAspSerLeuIleProAsnLeuGlyArgTyrIleThrGluGlnProMetAlaPheCysG1 380
Db 977 GGAT-----ATTCCAGGTGTGGCGCAACCTCTCTGACCAACCCCGAGACCATATTAT 1030
Qy 380 n-----IleValLeuArgGlnGluPheValAspSer----- 390
Db 1031 GTGGGAGCTCAACAAGCCCGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1090
Qy 391 ----ValArgAspAspProTyrGlyLeuProTrpTrpLysGluAlaValAlaGlnHisI1 409
Db 1091 TTTCTCTGCGACGAGAGCC-----AGAACGC 1117
Qy 409 eAlaLysAsnProThrAspAla-----LeuProIleProPheAr 422
Db 1118 TGTCTGTACGATGGCGATGCTGTGTATGTCATGTCATGTCATGTCATGTCATGTCATGTC 1177
Qy 422 gAspProGluProGlnValThrThrProPheThrGluGluHisProTrpHisThrGlnI1 442
Db 1178 CTTCAACACAGAGCGCTAGGCTATCTCT-----AT 1207
```

```
Qy 442 eHisArgAspAlaPheSerTyr----- 449
Db 1208 CATCAGGACGGTATGCGCTTTCGCATGACACCCCAACATCCCGCCCTCGCTCAGGTGG 1267
Qy 450 -----GlyAlaValGlyProGluValAspSerArgValIleValAspLeuAr 465
Db 1268 CGCATCTACTTGCATCGCCGACCTACTGTC-----AAGCTGCTCTCGATTTCGG 1321
Qy 465 gTrpPheGlyAlaThrAspProGluAlaAsnAsnLeuLeuValPheGlnAsnAspValG1 485
Db 1322 TTACTTC-----ACAGACCCTGAG----- 1340
Qy 485 nAspGlyTyrSerMetProGlnProThrPheArgTyrArgProSerThrAlaSerAsnVa 505
Db 1341 ----GGTTAGGAC-----GCTGCCACCTGGTCCATGGCAT 1372
Qy 505 lArgAlaArgLysMetMetAlaAspMetCysGluValAlaSerAsnLeuGlyGlyTyrLe 525
Db 1373 CAAGGCTGCCGAAAGATTGCGCAGCAGAGCCCTTCAAGGACTGGCTCAAGGAAGAAGT 1432
Qy 525 uProThrSerProGlnPheMetAspProGlyLeuAla----- 538
Db 1433 TGCCCTGTGTCGAAGATCCAAACGGATGAGGATCAGCGAATATGCTCGCGAGATTGC 1492
Qy 539 -----LeuHisLeuAlaGlyThrThrArgIleGlyPheAspLys-----AlaTh 553
Db 1493 CCACACAGTGTACCACCTGCGGTACCACTAAGATGGGTGAGCTGAGCGCGCATGAGAT 1552
Qy 553 rThrValAlaAspAsnAsnSerLeuValTrpAspPheAlaAsnLeuTyrValAlaGlyAs 573
Db 1553 GCGGTGTGTGACCCCGAGCTCAAGGTACGTGGAATCAGCAAGCTCCGCAATTGTTGATGC 1612
Qy 573 nGlyThrIleArgThrGlyPheGlyGluAsnProThrLeuThrSerMet 589
Db 1613 TGGTATC-----TTCCCGAAATGCCAACAATCAACCCCTATG 1649
RESULT 7
US-10-353-445-8
; Sequence 8, Application US/10353445
; Publication No. US20030119166A1
; GENERAL INFORMATION:
; APPLICANT: Baszczynski, Christopher L.
; APPLICANT: Lyznik, Leszek A.
; APPLICANT: Gordon-Kamm, William J.
; APPLICANT: Guan, Xueni
; APPLICANT: Rao, Guru
; APPLICANT: Tagliani, Laura A.
; TITLE OF INVENTION: A No. US20030119166A1 Method For The Integration Of Foreign
; FILE OF INVENTION: Eukaryotic Genomes
; FILE REFERENCE: 5718-66 (amended listing)
; CURRENT APPLICATION NUMBER: US/10/353,445
; CURRENT FILING DATE: 2003-01-29
; PRIOR APPLICATION NUMBER: US/09/193,503B
; PRIOR FILING DATE: 1998-11-17
; PRIOR APPLICATION NUMBER: 60/099,435
; PRIOR FILING DATE: 1998-09-08
; PRIOR APPLICATION NUMBER: 60/056,627
; PRIOR FILING DATE: 1997-11-18
; PRIOR APPLICATION NUMBER: 60/065,613
; PRIOR FILING DATE: 1997-11-18
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 2346
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Sequence
; OTHER INFORMATION: encoding a FLPm:Cre polypeptide, FLP from
; OTHER INFORMATION: Saccharomyces (maize preferred codons), and Cre
; OTHER INFORMATION: from Bacteriophage P1
; FEATURE:
```

```

; NAME/KEY: CDS
; LOCATION: (1)..(2346)
US-10-353-445-8

```

Alignment Scores:

Pred. NO.:	0.155	Length:	2346
Score:	108.50	Matches:	112
Percent Similarity:	32.98%	Conservative:	73
Best Local Similarity:	19.96%	Mismatches:	212
Query Match:	3.30%	Indels:	165
DB:	14	Gaps:	27

US-09-856-327-2 (1-618) x US-10-353-445-8 (1-2346)

Qy	129	ProGlyAlaTrpSerAlaProGlySerSerAlaIleSer-----AsnGly	144
Db	275	CCCTCAAGAAGCTCATCCCGCTGGAGGTTCACCATCATCCCTACTACGCCGCAAGC	334
Qy	145	LysAsnPro-----HisGlnArgGluPheGluAsnLeuSerAlaGlu	158
Db	335	ACCAGTCCGACATCACCAGCATCGTGTGTCATCCCTCCAGCTTCAGTTTCGAGTCTCCGAGG	394
Qy	159	AlaValThrArgGlyValGlyGlyMetSerThrHisTrpThrCys-----SerThr	175
Db	395	AGGCTCAAGGGCAACT-----CCCACTCCAAGAAGATGCTGAAGGCCCTCTCTCT	445
Qy	176	ProArgIleHisPro-----PrometGluSerLeu	185
Db	446	CCGAGGGCGAGTCCATCTGGGAGATCACCAGAGAAGATCCTCAACTCTCTCGAGTACACCT	505
Qy	186	ProGlyIleGlyArgProLys-----LeuSerAsnAspPro-----	197
Db	506	CCAGGTTCACTAAGACCAAGACCCCTTACCAGTTCTCTTCCCTCGCCACCTTCATCAACT	565
Qy	198	-----AlaGluAspAspLysGlyTrpAsnGluLeuTyrSerGluAlaGluArgLeu	214
Db	566	CGCGCAGGTTCCTCAGACATCAAGAACGTGACCCCAAGTCTCTTCAAGCTGCTGCAGAAC	625
Qy	215	Ile-----GlyThrSerThr-LysGluPheAspGluSerIleArg-----	227
Db	626	AGTACCTCGCGTGCATCATCCAGTCCCTCGTACCAGAGACCAAGACCTCGTGTCCAGGC	685
Qy	228	-HisThrLeuValLeuArgSerLeuGlnAspAlaTyrLysAspArgGlnArgIlePheAr	247
Db	686	ACATCTACTTCTTCTCGCTCGCGCAGGAT-----CG	718
Qy	247	q-ProLeuProLeuAlaCysHisArgLeuLysAsnAlaProGluTyrValGluTrpHis	267
Db	719	ACCCCTCGTGTACCTCGACGAGTTCCTCAGAACTCAGAGCCCGTCTCAAG-----	771
Qy	267	erAlaGluAsnLeuPheHisSerIleTyrAsnAspAspLysGlnLysLysLeuPheThrL	287
Db	772	-----AGGGTGAACAGGACCGGCAACTCCTCTCCAACAGCAGGAGTACCAGC	820
Qy	287	euLeuThrAsnHisArgCysThrArgLeuAlaLeuThrGlyGlyTyrGluLysLysIleG	307
Db	821	TCCTCAAGGACAC-----CTCGTGAGTCTCAACAAAGGCCCTCA	862
Qy	307	lyAlaAlaGluValArgAsnLeuLeuAlaThrArgAsn---ProSerSerGlnLeuAsps	326
Db	863	AGAAGACGCCCTACTTCCATCTTCGCCATCAAGACGGCCCAAGTCCCATCCGATCGGA	922
Qy	326	erTyrIleMetAla-----LysValTyrValLeuAlaSerGlyAlaIleG	341
Db	923	GGCACCCTCATGACCTCTCTCTCTCAATGAAGGGCCTCACCGAGCTCACCACACGTGTG	982
Qy	341	lyAsn-----ProGlnIleLeuTyrAsnSerGlyPheS	352
Db	983	GCAACTGGTCGACAGAGGGCTCCGCCCTGGCGAGGACCACTCAACCCAC-----	1035
Qy	352	erGlyLeuGlnValThrProArgAsnAspSerLeuIleProAsnLeuGlyArgTyrIleT	372
Db	1036	-----CAGATCCCGCCATCCCGACCACTACTTCGCCCTCTGCTCAAGGTACTACG	1087

Qy	372	hrGluGlnProMetAlaPheCysGlnIleValLeuArgGlnGlu	-----	386
		: : : :	: :	
Db	1088	CCTACGACCCCATCTCCAAGGAGATGATCGCCCTCAAGGACGAGACTAAACCCATCGAGG	1147	
Qy	387	-----PheValAspSerValArgAspAspProTyrGly	-----LeuProThrpt	401
		: : : : : : : : :		
Db	1148	AGTGGCAGCACATCGACGAGCTCAAGGGCTCCGCCGAGGGCTCCATCAGGTACCCCGCT	1207	
Qy	401	rplysGluAlaValAlaGlnHisIle	-----	409
		: : : : : : :		
Db	1208	GGAACGCATCATCTCCAGGAGGTGTCGCAGTACCTCTCCTCCTACATCAACAGGAGGA	1267	
Qy	410	-----AlaLysAsnProThr	-----	414
		: : : : : : :		
Db	1268	TCGGTGGCGCAGCGGTGGCGGCTCCCGCGGTGGCTCGGATCCAACCATGCCCAATTAC	1327	
Qy	415	-----AspAlaLeuProIlePro	-----	420
		: : : : : :		
Db	1328	TGACCGTACCAAAAATTTCCTGCATATCCGGTTCGATGTCGAACGAGTGTGAGGTTCCGA	1387	
Qy	421	-----PheArgAspProGluProGlnValThrProPheThrGluG	435	
		: : : : : :		
Db	1388	AGAACCTGATGGACATGTTCAGGATCGCCAGCGG	-----TTTTCTG	1429
Qy	435	luHisProTrpHisThrGlnIleHisArgAspAlaPheSerTyrGlyAlaValGlyProG	455	
Db	1430	AGCATCTCGAAAATGCTTCTG	-----TCCGTTTCCGCTCGTGGCGCATGGTGCA	1483
Qy	455	luValAspSerArgValIleValAspLeuArgTrpPheGlyAlaThrAspProGluAla	475	
		: : : : : : : :		
Db	1484	AGTTGAATAACCGG	-----AAATGGTTTCCCGCAGAACCTCAAGATGTC	1528
Qy	475	snAsnLeuValPheGlnAsnAspValGlnAspGlyTyrSerMetProGlnProThr	495	
		: : : : : : : :		
Db	1529	CGCATTTATCTTCATATCT	-----TCAGGCCGCGGCTGCGAGTAAACATATCCAGCACAT	1587
Qy	495	heArgTyrArgProSerThrAlaSerAsnValArgAlaArgLysMetMetAlaAspMetC	515	
		: : : : : : : :		
Db	1588	TTGGCGCAGCTAAACATGCTTCATCGTCGGTCCGGTCCGCGGTGCCACGCAAGTGACAGCAAT	1647	
Qy	515	ysGluValAlaSerAsnLeuGlyGlyTyrLeuProThrSerPro	-----ProGlnPheMetAsp	534
Db	1648	GC-----TGTTTCACTGGTTAT	-----GCGCGGGATTCGGAAGAAAACGTTGAT	1692
Qy	535	ProGly	-----LeuAlaLeuHisLeuAlaGlyThrThrArgIleGlyPheAspLysAlaThr	553
Db	1693	GCCGGTGAACGTGCAAAACAGGCTCTAGCGTTCGACGCATGATTTCCACCAGGTTCTG	1752	
Qy	554	ThrValAlaAspAsnAsnSerLeuValTrpAspPheAlaAsnLeuTyrValAlaGly	572	
Db	1753	TCATCTATGAAATACGATCGCTGCCAGGATATAGTAATCTGCGCATTTCTGGG	1809	
		: : : : : : : :		

RESULT 8

US-10-419-969-6
; Sequence 6, Application US/10419969
; Publication No. US20030175893A1

GENERAL INFORMATION:

APPLICANT: Agricultural University Wageningen
TITLE OF INVENTION: A novel method to isolate mutants and to clone the complementing gene

NUMBER OF SEQUENCES: 10

STREET: Costerweg 50

CITY: Wageningen

COUNTRY: The Netherlands

ZIP: 6701 BH

COMPUTER READABLE FORM:

```

;
MEDIUM TYPE: Floppy disk

```

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS

; SOFTWARE: PatentIn Release

;
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/1

/ FILING DATE: 22-Apr-2003
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US/08/981,729
/ FILING DATE: <Unknown>
/ INFORMATION FOR SEQ ID NO: 6:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 3026 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: double
/ TOPOLOGY: linear
/ MOLECULE TYPE: DNA (genomic)
/ HYPOTHETICAL: NO
/ ANTI-SENSE: NO
/ ORIGINAL SOURCE:
/ ORGANISM: Aspergillus niger
/ STRAIN: N400 (CBS 120.49)
/ FEATURE:
/ NAME/KEY: TATA_signal
/ LOCATION: 643..648
/ NAME/KEY: CDS
/ LOCATION: 724..2538
/ OTHER INFORMATION: /EC_number= 1.1.3.4.
/ /product= "glucose oxidase"
/ /gene= "goxC"
/ FEATURE:
/ NAME/KEY: mat_peptide
/ LOCATION: 790..2538
/ FEATURE:
/ NAME/KEY: sig_peptide
/ LOCATION: 724..790
/ SEQUENCE DESCRIPTION: SEQ ID NO: 6:
/ US-10-419-969-6

Alignment Scores:

Pred. No.: 0.242 Length: 3026
Score: 108.50 Matches: 136
Percent Similarity: 35.02% Conservative: 86
Best Local Similarity: 21.45% Mismatches: 233
Query Match: 3.30% Indels: 180
DB: 12 Gaps: 32

US-09-856-327-2 (1-618) x US-10-419-969-6 (1-3026)

```
QY 3 LeuSerThrGluGlnMetLeuArgAspTyrProArgSerMetGlnIleAsnGlyGlnIle 22
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 754 GTCCTCGCTGCGGCCCTGCCACACTACATCAGGAGCAATGGCATTAAGCCAGCCCTC 813
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 23 -----ProLysAsnAlaIleHisGluThrTyrGlyAsnAspGlyValAspValPhe 39
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 814 CTGACTGATCCCAAGAGTCTCCGGCGCGCAGC-----GTCGACTACATC 858
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 40 IleAlaGlySerGlyProIleGly---AlaThrTyrAlaLysLeuLysValGluAlaGly 58
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 859 ATCGTGTGGAGGTCTGACTGACTCACCGCGTCTGCTCAGGAGAACCCCAAC 918
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 59 LeuArgValMetValGluIleGlyAlaAlaAspSerPheTyrAlaValAsnAlaGlu 78
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 919 ATCAGTGTCTGTCATCAAGAGTGGCTCCTACGAGTCG-----GAC 960
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 79 GluGlyThrAlaValProTyrValProGlyTyrHisLysLysAsnGluIleGluPheGln 98
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 961 AGAGGT-----CCTATCATTTAGGACCTGAACGCTACGGCGCAGCATC---TTTGGC 1008
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 99 LysAspIleAspArgPheValAsnValIleLysGlyAlaLeuGlnGlnValSerValPro 118
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1009 AGCAGTGTAGACAGCCCTACGAGACCGGTGGAGCTCGCTTACCACAAATCAAAACCGCGCTG 1068
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 119 ValArgAsnGlnAsnValProThrLeuAspProGlyAlaTyrSerAlaProProGlySer 138
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1069 ATCCGCTCGGAAAT-----GATCTCGGTGGC 1095
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 139 SerAlaIleSerAsnGly-----LysAsnProHisGlnArgGluPheGluAsnLeu 155
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```

    ||| ... ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1096 TCTACTCTAGTGAATGGTGGCACCTCGGACTCCGCCCCACAGGCACAGGTTGCAC----- 1149
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 156 SerAlaGluAlaValThrArgGlyValGly-GlyMetSerThrHisThrCysSerTh 175
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1150 TCTTGGGAGACTGTCTTTGGAATAGGCTGGAACTGGGACAATGTG----- 1198
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 175 rProArgIleHisProProMetGluSerLeuProGlyIle-GlyArgProLysLeuSerA 195
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1199 -----CCGCTACTCTCCAGGCTGAGCGTCCGCGCACCAAA----- 1238
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 195 snAspProAlaGluAspAspLysGluTrpAsnGluLeuTyr-SerGluAlaGluArgLeu 214
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1239 -----TGCCAAACAGATCGTCTGCGCCACTACTTCAAGCATCTCTGCCATGGTGT 1290
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 215 IleGlyThr---SerThrLysGluPheAspGluSerIleArgHisThrLeuValLeuArg 233
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1291 AATGGTACTGTCCATGCCGACCCCGCGACACCGCGGATGAGTATTTCTCCCATCGTCAAG 1350
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 234 SerLeuGlnAspAlaTyrLysAspArgGlnArgIlePheArgProLeuProLeuAlaCys 253
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1351 GCTCTCATGAGCGCTCGAAGACCG----- 1377
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 254 HisArgLeuLysAsnAlaProGluTyrValGluTyrHisSerAlaGlu----- 269
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1378 -----GGCGTTCCCAAGAAAGACTTCGGATGCGGTACGCCCATGGTGTG 1425
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 270 AsnLeuPheHisSerIleTyrAsnAspAspLysGlnLysLysLeuPheThr----- 286
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1426 TCCATGTTCCCAACACACTTGCACGAAGACCAAGTCCGATCCGATCCGCTCGCAATGG 1485
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 287 LeuLeuThrAsnHisArgCysThrArgLeu---AlaLeuThrGlyGlyTyrGluLysLys 305
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1486 CTACTTCCCACTACCAAGCTCCCAACCTGCAAGTCTGACCGGACAGTATGTTGTAAG 1545
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 306 IleGlyAlaAlaGluValArgAsnLeuAlaThrArgAsnProSerSerGlnLeuAsp 325
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1546 GTGCTCTTAGTCAG-----AACGGCACCAACCTCGTCCGCTGGGTGAATTCGCGC 1599
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 326 SerTyr-----IleMetAlaLysValTyrVal---LeuAlaSerGly 338
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1600 ACCCAGAGGGCAACCCCAACAGTTTACGTTAAGCAGCAGGTCTCTCTGCGCGGGCGC 1659
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 339 AlaIleGlyAsnProGlnIleLeuTyrAsnSerGlyPheSerGlyLeuGlnValThrPro 358
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1660 TCCGCTGTCTCTCCCAATCTCTCAATATTCCGATATC---GGAATGAG----- 1707
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 359 ArgAsnAspSerLeuIleProAsnLeuGlyArgTyrIleThrGluGlnProMetAlaPhe 378
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1708 -----TCCATCTGGAGCCCTTGGT----- 1728
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 379 CysGlnIleValLeuArgGlnGluPheValAspSerValArgAspProTyrGlyLeu 398
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1729 -----ATCGACACCGTCTGTTGACTGCTGCCCTCGGCTGCTG 1761
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 399 ProTrpTrpLysGluAlaValAlaGlnHisIleAla-LysAsnProThrAspAlaLeuPr 418
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1762 AACCTGCAGGACCAACCCCGCT-----ACCGTCCGCTCCCGCATCACTCTGCT 1812
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 418 oIleProPheArgAsp-----ProGluProGlnValThrThrProPheThrGluG 435
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1813 GGTGAGGACAGGAGCGCGCTTGGTTGCGCACCTTCAACGAGACCTTTGGTGACTAT 1872
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 435 uHisProTrpHisThrGlnIleHisArgAspAlaPheSerTyrGlyAlaValGlyProG 455
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1873 TCCGAAAGGACACAGAGCTGCTCAACACCAAGCTGGAGCAGTGGCGGACGCGCG-- 1930
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 455 uValAspSerArgValIleValAspLeuArgTrpPheGlyAlaThrAspProGluAlaAs 475
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1931 -----TCGCCGCTGGCGGA 1944
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 475 nAsnLeuLeuValPheGlnAsnAspValGlnAspGlyTyrSerMetProGlnProThrPh 495
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

1945	TtCC-----	-----ACAAACACCACCGCTTGCTC	1968
Qy	495 eArgTyArgProSerThrAlaSerAsnVal-	-----ArgAlaArgLysMe	510
	::: :::	:::	
Db	1969 ATCCAGTACGAGAACTACCGCGACTGGATTGTC	CAACCACAAACGTCGCGTACTCGGA	2028
Qy	510 tMetAlaAspMetCysGlu--ValAlaSerAsnLeu	GlyGlyTyrLeuPro-----	526
	::: :: ::		
Db	2029 TTCTCTCGACACTGCCGGAGTAGGCAGCTTCGATGT	GTGGGACCTTCTGCCCTTCA	2088
Qy	527 -----ThrSerProGlnPheMetAspProGlyLeu	AlaLeuHi 540	
Db	2089 GGATACGTTACATCTCGACAAAGGACCCCTACTTC	ACCACTTCGCGCTACGACCCCTCAG	2148
Qy	540 sLeuAlaGlyThr-----ThrArgIleGlyPheAsp	LysAlaThrThrValAlaAspAs	558
	::: ::		
Db	2149 TACTTCTCAACGAGCTGGACCTG-CTCGGTGAGGTG	CGCGCTACTCAACTGGCCGCGCA	2207
Qy	558 nAsnSerLeuValTrpAspPheAlaAsnLeuTyrVal	AlaGlyAsnGlyThrIleArgth	578
		:::	
Db	2208 C-----ATCTCCAACCTCCGGTCCCATGCAGAC	2234	
Qy	578 rGlyPhe-----GlyGluAsn	583	
Db	2235 CTACTTCGCTGGGAGACTATCCCGGTGATAAC	2268	

[illegible]

```
QY 321 SerSerGlnLeuAspSerTyrIleMetAlaLysValTyrValLeuAlaSerGlyAlaIle 340
Db 1717 -----AAGTTCGGCGGTGGTGGTGGACGCCCAAAATGTCTGCGCGTACCTTG 1664
QY 341 -----GlyAsnProGlnIleLeuTyrAsnSerGlyPheSerGlyLeuGlnValThr 357
Db 1663 CGAGTGGCACAAAGCCCGAGCTCTCAA-----CCCGGTCTCAAAACAGGAC 1616
QY 358 ProArgAsnAspSerLeuIleProAsnLeuGlyArg-----TyrIleThrGlu 373
Db 1615 CCGCGAAAC-----AAGCGCGTCAAGTCCAGCACGCCCATCTGCGCGAT 1571
QY 374 GlnProMetAlaPheCysGlnIleValLeuArgGlnGluPheValAspSerValArgAsp 393
Db 1570 GTTCCACGTCACATCCGGGATCACAGTCAGTCGAGCGTTGTATG-----CTCCGCGT 1517
QY 394 AspProTyrGlyLeu-----ProTrpTrpLysGluAlaValAla 406
Db 1516 CATCCCGCAGCGATTGTTCGCTAACACAGCGTCTACCCGGCTTTCGAGATAGAAGC 1457
QY 407 Gln-----HisIleAlaLysAsnProThrAspAlaLeuProIleProPhe 421
Db 1456 CAGGCTCTGCTGCTTCACACAGAACGCCAGCACCTCTGCAACAGCCCATGAACAG 1397
QY 422 ArgAspProGluProGlnValThrProPheThrGluGluHisProTrpHisThrGln 441
Db 1396 CAGAACAGGATACCTCTGTTTCGACCCGAGGATTCGGCCTTCTCCGCCCAATAACAC 1337
QY 442 IleHisArgAspAlaPheSerTyrGlyAlaValGlyProGluVal-----AspSer 458
Db 1336 GTCCGG-----GCCGGGGCCCTCTCTCTCTTCCGACAAACGG 1298
QY 459 ArgValIleValAspLeuArgTrpPheGlyAlaThrAspProGluAlaAsnAsnLeu 478
Db 1297 CGGTACTCTCGGGGTAA-----AACATGCCAACAGTAAATATAAGCC 1250
QY 479 valPheGlnAsnAspValGlnAspGlyTyrSerMetProGlnPro-----493
Db 1249 ATCCCTCGTGTGAGCTTGAACGCTTCGACATCGAGCCCTTACCCGCGCGGTAAATACCT 1190
QY 494 -----ThrPheArgTyrArgProSerThrAlaSerAsnValArgAlaArgLysMet 510
Db 1189 AACATCACAAACGAGCAAGTCGGCGCCGCTCCCAACGGGGACACTATCCGGCGCCTC 1130
QY 511 MetAlaAspMetCysGluValAlaSerAsnLeuGlyTyrLeuProThrSerProPro 530
Db 1129 CTGCTC-----CACATCCCGCTTCGCGCTCGCGCGCTCCACCT 1091
QY 531 Gln-----PheMetAspProGlyLeuAlaLeuHisLeuAlaGly 543
Db 1090 TCTAAGCTCTCTCTCGACTCCACTGCTCGCTCGCGCTCTCTCGGC 1046

RESULT 10
US-10-128-714-5075/C
; Sequence 5075, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wenqi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroszhin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; FILE OF INVENTION: Methods of Use
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
```

```
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5075
; LENGTH: 3926
; TYPE: DNA
; ORGANISM: Aspergillus fumigatus
; US-10-128-714-5075

Alignment Scores:
Pred. No.: 0.406 Length: 3926
Score: 108.00 Matches: 129
Percent Similarity: 33.44% Conservative: 84
Best Local Similarity: 20.25% Mismatches: 211
Query Match: 3.29% Indels: 213
DB: 14 Gaps: 36

US-09-856-327-2 (1-618) x US-10-128-714-5075 (1-3926)

QY 37 AspValPheIleAlaGlySerGlyProIleGlyAlaThrTyrAla-----51
Db 3196 AATACCTGGCTACCGGCTCTCTGGATATATGGGTGTACGTATGCTTTTAACTCCTGTAGC 3137
QY 52 -----LysLeuCys 54
Db 3136 AAGCGCAGGTAGTCTGGTCTGTGGTCGTGGTAGCGCCGTGTGATTCCATCTCTGT 3077
QY 55 ValGluAlaGlyLeuArgValValMetValGluIleGlyAlaAlaAspSerPhe-----72
Db 3076 GTGGGCTTGGTATCGAC---GGAATGATGAGACTGGGATAGCTATAAGTCTCATGAAT 3020
QY 73 ---TyrAlaValAsnAlaGlu-----GluGlyThrAlaValPro---TyrValProG 88
Db 3019 AGATACGCTGTACATATGCAATGTCTTAATGAACATCAAAGCTGCCAGAACAGTCATAT 2960
QY 88 lYTrpHisLysLysAsnGluIleGluPheGlnLysAspIleAspArgPheValAsnValI 108
Db 2959 CTCATCACCCTAGTCA-----TTCTGTAAGGAC-----CGTGTCAC 2924
QY 108 leLysGlyAlaLeuGlnGlnValSerVal-----117
Db 2923 TGAACACACCTTTGGAAACCCGCGCACACTTCCGCGCGCTCGCATGGTCTGTATAT 2864
QY 118 -----ProValArgAsnGlnA 123
Db 2863 AACATCGCGGACCTCTTTGACCCACTTGTCCACCGCATCCATGGCCACAGCATCTAG 2804
QY 123 snValProThrLeuAspProGlyAlaTrpSerAlaProProGlySerSerAlaIle----141
Db 2803 ATGCTCATACTCTTCAA---TCACITTTGAATGACACCATCTGCACATGGCGCTCGCTCC 2747
QY 142 -----SerAsnGlyLysAsnProHisGlnArgGluPheGluAsnL 155
Db 2746 GCTGTGCAACCGCTGGAGGAGAGCGCGCGCTCGACCA-----GCGCATC 2702
QY 155 euSerAlaGlu-----AlaValThrArgGlyValGlyGlyMetSerThrHisTrpThrC 173
Db 2701 TGACCCCGCAACCCACAGCAGCAAAACGGGGCGTCTCGCGCCCGTACACAGCGGTGTACGT 2642
QY 173 ysSerThrProArg-----IleHisProPro-----181
Db 2641 GCGCGAGCCCGCTGCTCTCGCTCGCGCTGCAGACCGACATCTCTCTCCATCTCGCGGAG 2582
QY 182 -----MetGluSerLeuProGly---IleGlyArgProLysLeuSerAsnAspProAla-198
Db 2581 GGTCTCTTCCATCGTGGCCAGGATGACCGCTCGCTCGCGAAAGACAGCCCTGACCCAGCCA 2522
QY 199 GluAspAspLys-----GluTrpAsnGluLeuTyrSerGluAlaGluArg 213
Db 199 GluAspAspLys-----GluTrpAsnGluLeuTyrSerGluAlaGluArg 213
```


Db 2521 CCACCGATCGTCTCGCTGCTATACATACATGCGCGAATTG-----GAACAT 2474
QY 214 LeuileGlyThrSerThrLysGluPheAsp-----GluserIleArgHisThr 229
Db 2473 GCGATCGCGGAGTCGCGATCCACGCGACGTCGCTCCAGCGAACAGGTGCGCGAACAC 2414
QY 230 LeuVal-----LeuArgSerLeuGlnAspAlaTyrIlyAsp 241
Db 2413 ATGATACCCCAACGCCCGTAGATCGCGGATGGAGAAAGCGGTGGAGCGTTATCATCAT 2354
QY 242 ArgGlnArgIlePheArgProLeuAlaCysHisArgLeuLysAsnAlaProGlu 261
Db 2353 GGGGATGAA-----TCGTGGATGCGCGAAATCGTCCGGAA 2318
QY 262 TyrVal---GluTrpHisSerAlaGluAsnLeuPheHisSerIleTyrAsnAspAspLys 280
Db 2317 TATGTCAGCGGTATACATCGCATGAACCGAAGTACACCTCTC-----GATCAA 2267
QY 281 GlnLysLysLeuPheThrLeuLeuThrAsnHisArgCysThrArgLeuAlaLeuThrGly 300
Db 2266 CGGCCC----- 2261
QY 301 GlyTyrGluLysLysIleGlyAlaAlaGluValArgAsnLeuLeuAlaThrArgAsnPro 320
Db 2260 GGCATACACGCCCGGTGCGAGGCGACAGAATACAGATATCTTCTCGCC----- 2213
QY 321 SerSerGlnLeuAspSerTyrIleMetAlaLysValTyrValLeuAlaSerGlyAlaIle 340
Db 2212 -----AGTTCCGGCGGTGGTCTGGACAGCGCCACAATGTCGCGGGTACCTTG 2159
QY 341 -----GlyAsnProGlnIleLeuTyrAsnSerGlyPheSerGlyLeuGlnValThr 357
Db 2158 CGAGTGGCACACAAGCCCGCTCTCAA-----CCGGTCTCAACAGGAC 2111
QY 358 ProArgAsnAspSerLeuIleProAsnLeuGlyArg-----TyrIleThrGlu 373
Db 2110 CCGGGAAC-----AAGCGCGTCAAGTCCAGCACGCGCCATCTCGCGGAT 2066
QY 374 GlnProMetAlaPheCysGlnIleValLeuArgGlnGluPheValAspSerValArgAsp 393
Db 2065 GTTCCACGTCCACATCCCGGATCACCAGTCGACAGCGTTGTATG-----CTCGGCGT 2012
QY 394 AspProTyrGlyLeu-----ProTyrTrpLysGluAlaValAla 406
Db 2011 CATCCGCGAGGATGTTGCCCTAACACAGCGTCGTACCCGCTTGGCAGAGATAGAAAGC 1952
QY 407 Gln-----HisIleAlaLysAsnProThrAspAlaLeuProIleProPhe 421
Db 1951 CAGGCTGCTGCTGCTTTCACACAGACGACACCACTCTGCAACAGCCCATGAACACAG 1892
QY 422 ArgAspProGluProGlnValThrThrProPheThrGluGluHisProTyrHisThrGln 441
Db 1891 CAGAACAGGATACCTCTGTTTCCACCGAGGATTCGGCCCTTCCGCCCATTAACAC 1832
QY 442 IleHisArgAspAlaPheSerTyrGlyAlaValGluProGluVal-----AspSer 458
Db 1831 GTCCGG-----GCCCGGGGCCCTCTCTCTCTTCCGACAAAGG 1793
QY 459 ArgValIleValAspLeuArgTrpPheGlyAlaThrAspProGluAlaAsnLeuLeu 478
Db 1792 CGTGATACCTTCGCGGTTTAAA-----AACATGCCACACACGCTAATAATAAGGCC 1745
QY 479 ValPheGlnAsnAspValGlnAspGlyTyrSerMetProGlnPro----- 493
Db 1744 ATCTCCGCTCGGACTTGGACGTCTCGACATCGAGCCCTACCGCGCGCGTAAATACCT 1685
QY 494 -----ThrPheArgTyrArgProSerThrAlaSerAsnValArgAlaArgLysMet 510
Db 1684 AACATCAACACGAGCAAGTCGCGCCGCCCTCCAAAGCGCGACACTCATCCGCGCGCTC 1625
QY 511 MetAlaAspMetCysGluValAlaSerAsnLeuGlyGlyTyrLeuProThrSerProPro 530
Db 1624 CTCGTC-----CACATCCGCTTCCGCTTCCGCTTCCGCTTCCGCTTCCGCT 1586

QY 531 Gln-----PheMetAspProGlyLeuAlaLeuHisLeuAlaGly 543
Db 1585 TCTAAGCTCTCTCTCTCGACTCCACTGTCGCTCGCGCTCTCTCGGC 1541
RESULT 11
US-10-156-761-1/c
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIORITY FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1
Alignment Scores:
Pred. No.: 5.05e+04 Length: 9025608
Score: 107.50 Matches: 153
Percent Similarity: 31.99% Conservative: 94
Best Local Similarity: 19.82% Mismatches: 268
Query Match: 3.27% Indels: 260
DB: 14 Gaps: 34
US-09-856-327-2 (1-618) x US-10-156-761-1 (1-9025608)
QY 15 SerMetGlnIleAsnGlyGlnIleProLysAsnAlaIleHisGluThrTyrGlyAsnAsp 34
Db 1793303 TCCGTGAAGAACAACCGGACAGCTGTCCCGCAGAGGCGCTGCTCGCGGTTCGCG 1793244
QY 35 GlyValAspValPheIle---AlaGlySerGlyProIleGlyAlaThrTyrAlaLysLeu 53
Db 1793243 GGCCTCGCGCCCTGACGGGCTGCGGACGCGGCGGAGGCGGACACGAGGACCTC 1793184
QY 54 CysValGlu---AlaGlyLeuArgValValMetValGluIleGlyAlaAlaAspSerPhe 72
Db 1793183 TCCAAGAGCGGCGCGG-----CGCATGAAGAAGTACTACGGGCGGGGACCATTC 1793133
QY 73 TyrAlaValAsnAlaGluGluGlyThrAlaValProTyr---ValProGlyTyrHisLys 91
Db 1793132 AAGGCCCAAGGCGCTCACCTTCACGATCTGCACACAGCATGCCGACCTACCCGATG 1793073
QY 92 LysAsnGluIleGluPheGlnLysAspIleAspArgPheValAsnValIleLysGlyAla 111
Db 1793072 AAGCGGACTCGCTGTTCTTGAAGAGAGGTACCCGAGCGCAACCGCGGTC-----ACC 1793022
QY 112 LeuGlnGlnValSerValProValArgAsnGlnAsn----- 123
Db 1793021 CTCAAGCGGACCTCGGTCCCGCCGACGACTACGACAGAGAGCGAGCGCTGCTCATCGCG 1792962
QY 124 -----ValProThrLeuAspProGly-----Ala-TrpSerAl 134
Db 1792961 GCGGCGGACGCGCGCTTACTGATCCCAAGACGTACCGCGGCGGAGGACGCGGTTCTGTG 1792902


```
Db 1175461 GACATCGTGGCAGCGTCCCGCAAGGCGAAGCTCACCATGGATGACTACCGAGGGCGTTACC 1175402
Qy 399 oTrpTrpLysGluAlaValAlaGlnHisIleAlaLysAsnProThrAspAlaLeuProI 419
l::: |||||
Db 1175401 GTTT-----CCITGACCAACCCAGGTGGCATCGGT 1175372
Qy 419 ePro-----PheArgAspProGluProGlnValThrThrProPheThrGluGluHisPr 437
::: ||| |||||
Db 1175371 ACCGCCCACTGTTCACCGTCTAACCAAGGCGCAGGCACCATCATCGGTGTCGGTTCC 1175312
Qy 437 oTrpHisThrCln-----IleHisArgAspAlaPheSerTyrGlyAlaVala 452
||||| |||||
Db 1175311 ATGGATTACCCAGCAGATTCACAGGGCGTTTCAAGAACGCCCTTGCAGAGCTCGGGTT 1175252
Qy 452 lGly-----ProGluValAspSerArgValIleValAspLeuArgTr 466
: ||| ||| |||||
Db 1175251 GGCAAACTGTGCACCATCATCCCTCCACC-TACGATCACCGCTGATCCAG----- 1175204
Qy 466 pHeGlyAlaThrAspProGlu-----AlaAsnAsnLeuLeuVal----- 479
||||| ||| :|
Db 1175203 ---GGTCTGTCTCCGGTGAATTCCTCGCACCACATGTCTCGCTGTCTCACCGATGATC 1175148
Qy 480 -PheGlnAsnAspValGlnAspGlyTyrSerMetProGlnProThrPheArgTyrArg-- 498
||| :| |||||
Db 1175147 CTTCTGGGATGAGATCTTCGACCAATGAACGTTCTTACACCCCAATGCGTTGGGCACA 1175088
Qy 499 -----ProSerThrAlaSerAsnValArgAlaArgLysMetMetAlaAspMetCysG1 516
|||:| ||| ||| |||
Db 1175087 GGAGCTTCCAACACCGGTGTGTATAGAACACCCGGGTATG-----CAGCTCATGTA 1175034
Qy 516 uAlaSerAsnLeuGlyTyrLeu---ProThrSerProGlnPheMetAspPr 535
| :| ||| :|
Db 1175033 GGCATACCGTCCCGTGGACACCTCATCGGTGACCAACCACTTTTCATGGTTACGCC 1174974
Qy 535 oGlyLeuAlaLeuHisLeuAlaGlyThrArgIleGlyPheAspLysAlaThrThrVa 555
||||| :| :|
Db 1174973 TGGCATGCCA-----GTTCCAGACACCGCGACCTCGACAT 1174938
Qy 555 lAlaAspAsnAsnSerLeuValTrpAspPheAlaAsnLeuTyrValAlaGlyAsnGlyTh 575
: ||||| :|
Db 1174937 CGAGACCACAACTGACCATCTGGATCTGGACCGTACCTTCAACGTCTCGT----- 1174886
Qy 575 rIleArgThrGlyPheGlyGluAsnProThrLeuThrSerMetCysHisAlaIleLysSe 595
||||| |||||
Db 1174885 -----GGCTTCGCGCGCAAGAGACCATGACC-----CTGCGCGAGGTACTGTC 1174842
Qy 595 rAlaArgSerIleIleAsnThrLeuLysGlyGlyThrAsp 608
| ||||| |||||
Db 1174841 CCGCCTCGCGCTGCGGTACACCCCTCAAGGTGCGGTCCGAA 1174802
```

RESULT 14

```
US-10-156-761-6845
; Sequence 6845, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 6845
; LENGTH: 3306
```

```
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(3306)
US-10-156-761-6845

Alignment Scores:
Pred. No.: 0.523 Length: 3306
Score: 106.00 Matches: 121
Percent Similarity: 33.10% Conservative: 72
Best Local Similarity: 20.75% Mismatches: 202
Query Match: 3.23% Indels: 189
DB: 14 Gaps: 27
```

```
US-09-856-327-2 (1-618) x US-10-156-761-6845 (1-3306)

Qy 22 IleProLysAsnAlaIleHisGluThrTyrGlyAsnAspGlyValAspValPheIleAla 41
Db 1 GTGCCTAAGCGC-----ACCGATATCCAGTCCGTCTCGTCTGTCATC 39
Qy 42 GlySerGlyProIle-----GlyAlaThrTyr 50
Db 40 GGCTCCGGCCCGATCGTCTCGCCAGCCGCCGAGTTCGACTACTCCGGCACCCAGCGC 99
Qy 51 AlalysLeuCysValGluAlaGlyLeuArgValValMetValGluIleGly----- 67
Db 100 TGCCGCATCTCGCCGCCGAGGGCTCAGGGTCTATCTCGTGGTCAACTCCACCCGCGAGC 159
Qy 68 -----AlaAlaAspSerPheTyrAla----- 74
Db 160 ATCATGACCGACCCGGAGATCGCCGACGCCACCTACGTGAGCGCATCACCCCGAGTTC 219
Qy 75 -----ValAsnAlaGluGlu-----Gly 80
Db 220 GTCGAGAAGATCATCGCCAGGCGCGCGCGCGTGTGCCACCTCGGTGGTTCAG 279
Qy 81 ThrAlaValProTyrValProGlyTyrHisLysLysAsnGluIleGlu----- 96
Db 280 ACGCCCTGAAACACCGCCATCTCCATGCACGAGCAGGGTGTGCTGGAGAAGTACGGTTC 339
Qy 97 -----PheGlnLysAspIleAspArgPheValAsnVal 107
Db 340 GAGCTGATCGCGCCCAACGTGAGCGCATCAACAGGCGAGGACCGC-----GACCTG 393
Qy 108 IleLysGlyAlaLeuGlnGlnValSerValProValArgAsn-GlnAsnValPro----- 125
Db 394 TTCAAGGCGCTGTCGAGGCGCTCCGCGCAAGATCGGGCACGCGGAATCCGCCGCTCG 453
Qy 125 ----- 125
Db 454 GTCATCTGCCACTCCATGGAGCAGTGTCTCGAGGGCTCGAGACCTCGCGGGTTACCCC 513
Qy 126 -----ThrLeuAspProGlyAlaTrpSerAlaProProGlySerSer-AlaIleSerA 143
Db 514 GTCGTGCTCGCTCCTTCCATCCATGGCGCGCGCGCTCCGGCTCCGCGCAGCAGCAG 573
Qy 143 snGlyLysAsnProHisGlnArgGluPheGluAsnLeuSerAlaGluAlaValThrArgG 163
Db 574 GAGGAGCTGCGCCGCAT-----CGCG 594
Qy 163 lyValGlyGlyMetSerThrHisTrpThrCysSerThrProArgIleHisProMetG 183
Db 595 GGTACGG-----CCTGACGCTCTCCCGACCCACC 624
Qy 183 luSerLeuProGlyIleGlyArgProLysLeuSerAsnAspProAlaGluAspLysG 203
Db 625 GAGGTGCTCCGAGGAGTCCATCTCGGCTGGAAGAGTACGAGTACGAGTGGAGTGTATGCGC 684
Qy 203 luTrpAsnGluLeuTyrSerGluAlaGluArg-LeuIleGlyThrSerThrLysGluPhe 222
Db 685 GA-----CAAGAACGACAAACGTGCTGCTGCTCTCCATCGAAGACTTC 729
```

```
QY 223 AspGluSerIleArgHisThr-----LeuValLeuArgSerLeuGlnAspAlaTyr 239
Db 730 GACCCGATGGCGTGCACACCGGTACTCGATCACCGT-CGCGCCCTCGATGACGCT--- 785
QY 240 LysAspArgGlnArgIlePheArgProLeuAlaCysHisArgLeuLysAsnAla 259
Db 786 ---GACCGACCGCGAGTA-----CCAGCGCT---GCGCGACATCGGTATCGCATCAT 833
QY 260 ProGluTyrValGluTrpHisSerAlaGluAsnLeuPheHisSerIleTyrAsnAsp 279
Db 834 CGGTGAGTTCGTGTGCACAC-----CGCGCGTGCACATCCAGTTCGCGTGAACCC 887
QY 280 LysGlnLys-LysLeuPheThrLeuLeuThrAsnHisArgCysThrArgLeuAla----- 297
Db 888 CGACGACGCGCGATCATGTCTGAGATGAACCGCGCGTGTCCCGGTCTGTCGCGGT 947
QY 298 -----LeuThrGlyGlyTyrGluLysLysIleGlyAlaAla----- 309
Db 948 CGCCTCCAAGCGCGCGTTCGCGATCGCGAAGATCGCGGAGCTCGCGTTCGGCTA 1007
QY 310 -----GluValArgAsnLeuLeuAlaThrArgAsnProSerSer-----G1 323
Db 1008 CACGCTCGACGAGATCCCAACGACATCACCGAGAGACCCCGCGCTCTTCGACGCCGAC 1067
QY 323 nLeuAspSerTyrIleMetAlaLysValTyrValLeuAlaSerGlyAlaIleGlyAsnPr 343
Db 1068 GCTCGAC---TACGTGGTGTGAAGGCC----- 1094
QY 343 oGlnIleLeuTyrAsn-----SerGlyPheSerGlyLeuGlnValThrProArgAs 360
Db 1095 GCGCTTCGCTTTCGAGAAGTTCCCGAGCGCCGACTCGACCTTCACCCACCATGAAGTC 1154
QY 360 nAspSerLeuIleProAsnLeuGlyArgTyrIleThrGluGlnProMetAlaPheCysGl 380
Db 1155 GGTGCGGAGGCCATGGCGCATCGCGCGGAACTTCACCGAGGCG-----CTGCA 1202
QY 380 nIleValLeuArgGln-----GluPheValAspSerValAr 392
Db 1203 GAAGCGCTGCGGTCCTCGAGAAGAGGCTCGCAGATTCAGCTTCGTGCGGCGCGGG 1262
QY 392 qAspAspProTyrGlyLeuProTyrTrpLysGluAlaValAlaGlnHisIleAlaLysAs 412
Db 1263 CGAC-----AAGGCCCTTCCTTGGAGGAGCGCGTCCG 1295
QY 412 nProThrAsp-----AlaLeuProIleProPheArgAspProGluProGl 427
Db 1296 TCCCAACGACGCGGCATCAACTCGGTATGACGAGCCATCCGCGCGGCGGACGCCGGA 1355
QY 427 nValThrThrProPheThrGluGluHisProTyrThrHisThrGln-----IleHisAr 444
Db 1356 GGAGTCTTCGAGCGCAGCAGAGATGACCCCTGGTTCGTGACACAGCTTCTCTGATCAA 1415
QY 444 qAspAlaPheSerTyrGlyAlaValGlyProGluValAspSerArgValIleValAspLe 464
Db 1416 GGAATCGCGGACGAGTGGCGCGCGACAAAGCTGGACCGCGGAGCTGCTCGCGGAGGC 1475
QY 464 uArgTrpPheGlyAlaThrAspProGluAlaAsnAsnLeuLeuValPheGlnAsnAspVa 484
Db 1476 CAAGCGGACGCGCTTCTCGAGGTCCAGATCCCGAGATCCCGCGGCTCGCGGAGGAGCT 1535
QY 484 l 484
Db 1536 c 1536
RESULT 15
US-10-014-717-1/c
; Sequence 1, Application US/10014717
; Publication NO. US20020192778A1
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
```

```
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/10/014,717
; CURRENT FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US/09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-10-014-717-1

Alignment Scores:
Pred. No.: 49.6 Length: 68750
Score: 106.00 Matches: 106
Percent Similarity: 33.26% Conservative: 47
Best Local Similarity: 23.04% Mismatches: 184
Query Match: 3.23% Indels: 123
DB: 13 Gaps: 25

US-09-856-327-2 (1-618) x US-10-014-717-1 (1-68750)
QY 116 SerValProValArgAsnGlnAsnValProThrLeuAspProGly-----AlaTrp 132
Db 29586 TCGGTCTCTCGGACATAGCTCGGTGACGAGCATCTCGCCCTCGATCGGGTGGTGG 29527
QY 133 -----SerAlaPro-----ProGlySerSerAla 140
Db 29526 GSCGGGCTCGTTCGGGATCCAGGTCCAGAGCCGCCGCCCGCGGATGCTCGAGC 29467
QY 141 IleSerAsnGlyLysAsnProHisGlnArgGluPheGluAsnLeuSerAlaGluAlaVal 160
Db 29466 GCGGCGCCCGGCCCATCGCCCAT-----AACGCGCGCTGACAAAGGGCGG 29422
QY 161 ThrArgGlyValGlyGlyMetSerThrHisTrpThrCysSerThrProArgIleHis-Pr 180
Db 29421 ATCGCAGGC-----TCGTCCGCAACGATGATGCC 29392
QY 180 oProMetGluSerLeuProGlyIleGlyArgProLysLeuSerAsnAspProAlaGluAs 200
Db 29391 CCCC---GGGTCAACGACCCAGAGTCGGGCGCAACAGACACACGCGTCTCAGAAACCGAGCC 29335
QY 200 pAspLysGluTrpAsnGluLeuTyrSerGluAlaGluArgLeuIleGlyThrSerThrly 220
Db 29334 AGCCGA-----GCACCGCGCGG 29317
QY 220 sGluPheAspGluSerIleArgHisThrLeuValLeuArgSerLeuGlnAspAlaTyrIly 240
Db 29316 GTAGCAGCAGCGGTGCGTCCG-----CGATCTCATCATGATCCACGCCCTCC 29272
QY 240 sAspArgGlnArgIlePheArgProLeuPro-----LeuAlaCysHisArgLe 256
Db 29271 GCACCGCAGCAGCG-----CGTCCAGACCCCAAGTAGAGCACTACCTGCCAATCGCTT 29218
QY 256 uLysAsnAlaProGluTyrValGluTrpHisSerAlaGluAsnLeuPheHisSerIleTyr 276
Db 29217 C----- 29217
QY 276 rAsnAspAspLysGlnLysLysLeuPheThrLeuThrAsnHisArgCysThrArgLe 296
Db 29216 ---GACCGCGCGCAGCCTCGG-----TCACCAAGTTCGCGGTGCGGAT 29176
QY 296 uAlaLeuThrGlyGlyTyrGluLysLysIleGlyAlaAlaGluValArgAsnLeuLeuAl 316
Db 29175 GTCCTCTCGCGCGCATGGAGCAGCAGCATG-----GAAGTCCACGCTCGTCACAGC 29125
QY 316 aThrArgAsnProSerSerGlnLeuAspSerTyrIleMetAlaLysValTyrValLeuAl 336
Db 29124 GCTGACGCGACCGCTCTCGCGCACTCCACCTTAT-----CCGCCAATACC 29080
```

```
QY      336 aSerGlyAlaIleGlyAsnProGlnIleLeuTyrAsnSerglyPheSerglyLeuGlnVa 356
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
29079   AGCCAGCTCCCGCGCTCGCTCTCTGATT-----TCTGGA-----GGCTGCGAGT 29032

QY      356 lThrProArgAsnAspSerLeuIleProAsnLeuGly-----ArgTyrIleThrGl 373
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
29031   ATCTCCGGCCAGTCCACCGATAGAACACGAGCTTGGTTCGGATCCGCGCTCGTGGCGA 28972

QY      373 uGlnProMetAlaPheCysGlnIleValLeuArgGlnGluPheValAspSerValArgAs 393
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
28971   CGGCTGTTCAGGCTCGATCGGGCCAGTACCGCTCCCGCTGCCACGGATAGTTCGGCAGC 28912

QY      393 pAspProTyrGlyLeuProTyrTrpPlyGluAlaValAlaGlnHisIleAlaLysAsnPr 413
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
28911   GGCACCCCGCGCCAGCGTGGGAAGACGCCCGCCAGCTGACCGAGCGCGCGCGCGCC 28852

QY      413 oThrAspAlaLeuProIleProPheArgAspProGluPro-----GlnValTh 429
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
28851   CACA---GCCTCGCCAGCGCTCGAGCACCCCGCAGCTCTCGCGCCCGCGCGCAAC 28795

QY      429 rThrProPheThrGluGluHisProTyrHisThrGlnIleHisArgAspAlaPheSerTy 449
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
28794   GAGCGCAGCAGCGTCGGCTCCCTCCCGCAGCGAGGCTG---GCAACAGCCCGCAGCAGC 28738

QY      449 rGlyAlaValGlyProGluValAspSerArgValIleValAspLeuArgTrpPheGlyAl 469
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
28737   GTCGGCTTCGGGCCCA-----CTTCGACGAACGTGCCCGCACCGGCTTCGTGCAGC 28687

QY      469 aThrAspProGluAlaAsnAsnLeuLeuValPheGlnAsnAspValGlnAspGlyTyrSe 489
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
28686   GCCTTCACCCCGCTCCGCGA-----AG 28666

QY      489 rMetProGlnProThrPheArgTyrArgProSerThr---AlaSerAsnValArgAlaAr 508
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
28665   CGCACCGCTCCGCGA---CGTGGCGCACCCAGTACCCCGCGCGCTCAGCTCGTCCGCG 28609

QY      508 gLysMetMetAlaAspMetCysGluValAlaSerAsnLeu---GlyGlyTyrLeuProTh 527
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
28608   ACCACCTTCGCGCTCAGGTGCTCACCGAAGAGCTTGGCGCGCGGTACGTACCGCAGC 28549

QY      527 rSerProProGlnPheMetAspProGlyLeuAlaLeuHisLeuAlaGlyThrThrArg 546
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
28548   GCCGCCACCCGCGCAACTCTCCAGCATCGGTTCCTACGCGCGGAGTGGGACGCGT 28491
```

Search completed: September 24, 2003, 15:19:29

Job time : 23190 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 24, 2003, 05:57:01; Search time 104 seconds
(without alignments)
2622.834 Million cell updates/sec

Title: US-09-856-327-2
Perfect score: 3284
Sequence: 1 MSLSPEQMLRDPYRSMQING.....IINTLKGTGDKNTGHRNL 618

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters: -MODEL-frame+ p2n model -DEV-xlh

-Q-/cgn2_1/USPTO_spool/US09856327/runat_23092003_153420_15557/app_query.fasta_1.775
-DB-issued_Patents_NA -QFMT-fastap -SUFFIX-p2n.rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS-bits -SPART=1 -END=1 -MATRIX-blosum62 -TRANS-human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE-LOCAL -OUTFMT=ptto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER-US09856327@cgn1.1.56.0runat_23092003_153420_15557 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT_DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA: *
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq: *
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq: *
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq: *
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq: *
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq: *
6: /cgn2_6/ptodata/2/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1232.5	37.5	1869	1	US-08-734-925-1
2	1210	36.8	1869	3	US-09-305-381-1
3	1091	33.2	1701	3	US-09-023-731-2
4	198	6.0	4665	4	US-09-297-937C-12
5	191.5	5.8	1845	4	US-09-297-937C-8
c 6	148	4.5	4403765	3	US-09-103-840A-2
c 7	148	4.5	4411529	3	US-09-103-840A-1
c 8	146	4.4	1944	4	US-09-252-991A-9081
c 9	128.5	3.9	42225	4	US-08-311-731A-131
10	126	3.8	1611	4	US-09-252-991A-11302
11	124	3.8	2576	3	US-09-265-108-1
12	124	3.8	2576	3	US-09-479-264-1

13	122.5	3.7	2106	4	US-09-252-991A-11428
14	122.5	3.7	2400	3	US-08-930-001-1
15	122.5	3.7	3009	4	US-09-252-991A-11409
16	115.5	3.5	20235	1	US-07-642-734C-3
17	115.5	3.5	20235	3	US-08-439-009A-3
18	115	3.5	1758	4	US-09-252-991A-1644
19	114	3.5	4403765	3	US-09-103-840A-2
20	114	3.5	4411529	3	US-09-103-840A-1
c 21	113.5	3.5	1932	4	US-09-252-991A-7230
22	113.5	3.5	2898	4	US-09-252-991A-7467
23	109.5	3.3	1863	3	US-09-199-229-1
24	109.5	3.3	1863	3	US-09-443-087-1
25	109.5	3.3	1863	3	US-09-687-298-1
26	108.5	3.3	1848	3	US-08-333-802-1
27	108.5	3.3	2346	3	US-09-193-503B-8
28	108.5	3.3	2346	4	US-09-415-839-8
29	108.5	3.3	2788	4	US-09-347-878-45
30	108.5	3.3	3026	3	US-08-981-729-6
31	107.5	3.3	3396	3	US-08-974-549A-642
32	107	3.3	4725	4	US-09-410-551B-24
c 33	106.5	3.2	1230025	4	US-09-198-452A-1
34	106	3.2	1773	4	US-09-134-001C-1678
c 35	106	3.2	68750	3	US-09-335-409-1
c 36	106	3.2	68750	4	US-09-568-102-1
c 37	106	3.2	68750	4	US-09-567-969-1
c 38	106	3.2	68750	4	US-09-568-480-1
c 39	106	3.2	68750	4	US-09-568-486-1
c 40	106	3.2	68750	4	US-09-568-472-1
c 41	106	3.2	68750	4	US-09-567-899-1
c 42	105	3.2	1080	4	US-09-252-991A-11464
43	104.5	3.2	32798	4	US-09-604-694B-1
44	104.5	3.2	34303	2	US-08-735-609-4
45	104.5	3.2	34303	2	US-08-735-609-4

ALIGNMENTS

RESULT 1
US-08-734-925-1
; Sequence 1, Application US/08734925
; Patent No. 5712139
; GENERAL INFORMATION:
; APPLICANT: NISHIMURA, IKURO
; APPLICANT: OKADA, KIMI HARU
; APPLICANT: MINAMIHARA, TOMOYUKI
; APPLICANT: KAWAI, GENSIRO
; APPLICANT: KOYAMA, YASUJI
; APPLICANT: SUZUKI, MASARU
; TITLE OF INVENTION: NOVEL PYRANOSE OXIDASE, PYRANOSE OXIDASE
; TITLE OF INVENTION: GENE, NOVEL RECOMBINANT DNA AND PROCESS FOR PRODUCING
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBOLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/734,925
; FILING DATE: 22-OCT-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/568,428
; FILING DATE: 06-DEC-1995
; ATTORNEY/AGENT INFORMATION:


```
QY 171 TrpThrCysSerThrProArgIleHisProProMetGluSerLeuProGly----- 187
Db 379 TGGTCAGCGGTACATTTCCGCTGACCGTGAATGCTGCGCAGCAATATGAA 438
QY 188 -----IleGlyArgProLysLeuSerAsnAspProAlaGluAsp-----LysGlu 203
Db 439 GCGCGTTACGGCAAAACTTTATCCCGAAGGATGACGATTTCAGGATTTCCGGGTGACG 498
QY 204 TrpAsnGluLeu-----TyrSerGluAlaGluArgLeuIleGlyThrSer----- 218
Db 499 TATAACGAATGGAACCTTCTTCGATCAGCGCGAGAAAGTCTTTGGTACCTCGGGCAGT 558
QY 218 ----- 218
Db 559 GCCTGGACCATCAAGCAAGATGATCGCGAAGGAGAAAGCGCGCAACTTTTACGCGCCG 618
QY 219 -----ThrLysGluPheAspGluSerIle-ArgHisThrLeuValLeuArgSerLeuG1 236
Db 619 GACCGCTCCAGGACTTCCCGTGGCCGCGACAGAGCGGACTTACTCGGCGCAGCTGTTT 678
QY 236 nAspAlaTyrLysAspArgGlnArgIlePheArgProLeuProLeuAlaCysHisArgLe 256
Db 679 GCCAGCGCGCAGAGTCGGTGGGTATCATCCGTACGATA-----TGCCATCGGCC 729
QY 256 uLysAsnAlaProGluTyrValGluTrpHisSerAlaGlu----- 269
Db 730 -AACACCTCAGGTCCGTACCAACACACTACGGCGCACAGATGGCGCGTGCACACTCTG 788
QY 270 -----AsnLe 271
Db 789 CGGTATTGCGCGGTACCGCTGCTACATGTATTCAAAGCGTCGCTTAACGTGAACAT 848
QY 271 uPheHisSerIleTyrAsnAspLysLysLeuPheThrLeuLeuThrAsnHi 291
Db 849 CCGTCCGCGCTG-----CGTCAGGAGCCCAAGTTTCAGCTGCGTAAACACGC 896
QY 291 sArgCysThrArgLeuAlaLeuThrGlyGlyTyrGluLysLysIleGlyAla----- 308
Db 897 ATATGTGTTCGGGTCAATCTGACCGCGGACAAAAACGGCCACTGGCTGACCTATCT 956
QY 309 ----AlaGluValArgAsnLeuAlaThrArgAsnProSerSerGln-----LeuAs 325
Db 957 CGATGTTCAGGTGCGTGAAGTGTG-----CAGCCTGCGGATCTGTGTATCTGTCT 1007
QY 325 pSerTyrIleMetAlaLysValTyrValLeuAlaSerGlyAlaIleGlyAsnPro---G1 344
Db 1008 AGCGTTCAGTTCCACAACTGCACCTGATGCTGCTCGGTATCGCCAGCGATATAA 1067
QY 344 nIleLeuTyrAsnSerGlyPheSerGlyLeuGlnValThrProArgAsnAspSerLeu11 364
Db 1068 CCGGATCCTACGAAAGGTGTGTCGGCGCTAACTTCGCTTATCAGAACTCTCGACGCT 1127
QY 364 eProAsnLeu---GlyArgTyrIleThrGluGlnProMet-----AlaPheCysG1 380
Db 1128 GAAAGCGCTTTCGACAAACACCACTACACCGTTTATCGGTGCGGTGGCGGAGG 1187
QY 380 nIleValLeuArgGlnGluPheValAspSerValArgAspProTyrGlyLeu----- 398
Db 1188 GTGCGGTGGTGAATTTCAACCGCCGACAACTTCGACCGCGCCGCTACGCTTCGTCGG 1247
QY 399 -----ProTyrTrpLysGluAlaValAlaGlnHisIleAlaLysAsnProThrAspAl 416
Db 1248 TGGCTCGCCATTTCTGG-----GTGAACCGAGCGGGTACCAAAACCGGTTTCGG 1295
QY 416 aLeuProIleProPheArgAspProGluProGlnValThrThrProPheThrGluGluHi 436
Db 1296 TCTGCCGACCCCAAGGACGCGCGAAC----- 1323
QY 436 sProTyrHisThrGlnIleHisArgAspAlaPheSerTyrGlyAlaValGlyProGluVa 456
Db 1324 ----TGGGCGAGCGCATGGNAA-----GCCGCGGTGCGGATACCTA 1361
QY 456 lAspSerArgValIleValAspLeuArgTrpPheGlyAlaThrAspProGluAlaAsnAs 476
```

```
Db 1362 CAACCACCATATTTTCGATGGATGCCAC-----GGTGGCACCAAGTCATACCGCGCTAA 1415
QY 476 nLeuLeuValPheGlnAsnAspValGlnAspGlyTyrSerMetPro-----GlnProTh 494
Db 1416 CTACCTCGATCTCGATCCGAACTACAAATAAGTCTACGGCAGCGCTGCTGCGTATGAC 1475
QY 494 rPheArgTyrArgProSerThrAlaSerAsnValArg---AlaArgLysMetMetAlaAs 513
Db 1476 CTTTGATGCGCAGGAC-----AACGACATCAGGATGCGCAGTTTATGTCGCGAA 1526
QY 513 pMetCysGluVal-----AlaSerAsnLe 521
Db 1527 GATCGCGAAATCACCGAGGCCATGAATCCGAAGATGATCATCGCGCGCTAAGGGACC 1586
QY 521 uGlyGlyTyrLeuProThrSerProProGlnPheMetAspProGlyLeuAlaLeuHisLe 541
Db 1587 GGTACCATCTTCATACCAACCGTGTATCAA-----ACCAGCATAT 1628
QY 541 uAlaGlyThrThrArgIleGlyPheAspLysAlaThrValAlaAlaAspAsnSerLe 561
Db 1629 GAGCGCGGGCGCATCATCGGTGAAGATCCGAAACACGAGCAGTGAACCGTTATTGCA 1688
QY 561 uValTrpAspPheAlaAsnLeuTyrValAlaGlyAsnGlyThrIleArgThrGlyPheG1 581
Db 1689 GAGCTGGGATGTCGCAACGTGTTTGTGCGGGTGTCTCCGCGTTCGCCGAGGCTCTGG 1748
QY 581 yGluAsnProThr 585
Db 1749 CTACAACCCGACC 1761

RESULT 6
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Alignment Scores:
Pred. No.: 2.09 Length: 4403765
Score: 148.00 Matches: 144
Percent Similarity: 32.37% Conservative: 80
Best Local Similarity: 20.81% Mismatches: 206
Query Match: 4.51% Indels: 264
DB: 3 Gaps: 31

US-09-856-327-2 (1-618) x US-09-103-840A-2 (1-4403765)
QY 37 AspValPheIleAlaGlySerGlyProIleGlyAlaThrTyrAlaLysLeuCysValGlu 56
Db 3819533 GAGCTCTGTATTATTGATCGGTTTCGGGGGAGCGTCAACCGCTGCGGTCAGCGAA 3819474
QY 57 AlaGlyLeuArgValValMetValGluIleGlyAlaAlaAspSerPheTyrAlaValAsn 76
```

Db 3819473 AAGGGCTACCGGGTGGCGTCTCGAGCGCGG----- 3819441
Qy 77 AlaGluGluGlyThrAlaValProTyrValProGlyTyrHisLysLysAsnGluIleGlu 96
Db 3819440 -----CGCCGGTTCTCCGACGAGGAG 3819420
Qy 97 PheGlnLys-----AspIleAspArgPheValAsnValIleLysGlyAlaLeuGln 113
Db 3819419 TTCGCCAAGACGTCGTGGGACCTCGCAAGTTTCTC----- 3819384
Qy 114 GlnValSerValProValArgAsnGlnAsnValProThrLeuAspProGlyAlaTrpSer 133
Db 3819383 -----TGGGCG 3819378
Qy 134 AlaProProGlySerSerAlaIleSerAsnGlyLysAsnProHisGlnArgGluPheGlu 153
Db 3819377 CCGAGGTTGGCTCTATGGCATT-----CAACGTATCCACCGCTCGCAACGTG 3819327
Qy 154 AsnLeuSerAlaGluAlaValThrArgGlyValGlyGlyMetSerThrHisTrpThrCys 173
Db 3819326 ATGATCCTGGCGCGCGC-----GGGTGGCGGGGCTCCCTCAACTAC---GCC 3819279
Qy 174 SerThrProArgIleHisProProMetGluSerLeuProGlyIleGlyArgProLysLeu 193
Db 3819278 AACACGTGTACGTGCGCGCG----- 3819258
Qy 194 SerAsnAspProAlaGluAspLysGluTrpAsnGluLeu----- 207
Db 3819257 -----GAGCGGTCTTCGCGCACGACGAGTGTACACATCACCGACTGCGTGGCGAG 3819204
Qy 208 -----TyrSerGluAlaGluArgLeuIleGlyThrSerThrLys----- 220
Db 3819203 CTGATCGCGCACTACCAAGCAGCGCAGCGGATGTCGGGGT-GGTCAAAACCCGACCTT 3819145
Qy 221 -----GluPheAspGluSerIleArgHis----- 228
Db 3819144 CACCGACGCGCACCCGATCTGCAAGAGGTGCGCGACGAGATGGGTGTGCGCACACCTG 3819085
Qy 229 -----ThrLeuValLeuArgSerLeuGlnAspAlaTyrLysAsp 241
Db 3819084 GGTGCCGACCCGGTGTGAGTGTCTTCGGCCCCGCGCACCAAGACGCCCGCGCAAGAC 3819025
Qy 242 ArgGlnArg-----IlePheArgProLeuProLeuAlaCysHisArgLeuLysAsnAla--- 259
Db 3819024 CGTGCCCGATCCCTACTTCCGGCGTGGCGCGCGCGCACCGCTGCCTGGAATGCG 3818965
Qy 259 ----- 259
Db 3818964 CTGCTGTATGACGGCTGTCGCCACGCGTGCCAGAACACCTGTGTGAAAACTACCTTGG 3818905
Qy 260 ProGluTyrValGluTrpHisSerAlaGluAsnLeuPheHisSerIleTyrAsnAspAsp 279
Db 3818904 CCTCGCGAATCAGCTGCGCGCAAGTGTAT-----TCCGATGAC 3818866
Qy 280 -LysGlnLysLysLeuPheThrLeuLeuThrAsnHisArgCysThrArgLeuAlaLeuTh 299
Db 3818865 CACGGTGAAG----- 3818856
Qy 299 rGlyGlyTyrGluLysLysIle---GlyAlaAlaGluValArgAsnLeuLeuAlaThrAr 318
Db 3818855 -----GGGTTTGAAGCGGCTCGACGCGGTGTGGAGGTGCGCACAGTG----- 3818811
Qy 318 gAsnProSerSerGlnLeuAspSerTyrIle-----MetAlaLys 331
Db 3818810 -----CGAACCGCAGTGTGGCTCGCGCGGATCGGCGCACCTTCACCGCAAC 3818764
Qy 331 sValTyrValLeuAlaSerGlyAlaIleGlyAsnProGlnIleLeuTyr----- 347
Db 3818763 GCAGCTGTGTGGCGCGCGGCGCACCTGGGGGACCCAGCATCTGCTGTTCAGAGATGCGGGA 3818704
Qy 348 -----AsnSer--- 349
Db 3818703 TAGGGGTGCGTCCCTGGTCTCTCGAAGCGATGTTGGGTCTCTTACCCGACCAACTCAGA 3818644

Qy 350 -----GlyPheSerGlyLeuGlnValThrProArgAsnAsp----- 361
Db 3818643 ATCGATCGTCGGGGCCGCCACCTTGAAGGTCAATCCGACGCTGACCCACCGCGGT 3818584
Qy 362 -----SerLeuIleProAsnLeuGlyArgTyrIleThrGluGlnProMetAl 377
Db 3818583 GGCATCACGTCGTCGATTCCACCGACGCGCGCACCCACATC-----GAACCGGTGCG 3818530
Qy 377 aPheCysGlnIleValLeuArgGlnGluPheValAspSerValArgAspAsp----- 394
Db 3818529 CTACGGCAAGCTCCACCGCATGGGGTCTGCACACCTTGTATACCGACGCTCCGG 3818470
Qy 395 -ProTyrGly-----LeuProTrpTrpLysGluAlaValAlaGlnHisIleAlaLysAs 412
Db 3818469 TCCGACGGGCACCGCATGTCGCGCTGGAGGCAG---TTGCTGCAGACGCGCAGTCAGGA 3818413
Qy 412 nProThrAspAlaLeuProIleProPheArgAspProGluProGlnValThrThrProPh 432
Db 3818412 TCCG-----CGCGGCAC-CATCCGATGCTCAATCCCGCGC 3818378
Qy 432 eThrGluGluHisProTrpHisThrGlnIleHisArgAspAlaPheSerTyrGlyAlaVa 452
Db 3818377 AATCGAG-----CGAGCCACCGTATCGCGCTGTCATGACGACCC 3818336
Qy 452 lGlyPro-----GluVal-AspSerArgValIleValAspLeuArgTrpPheGlyA 469
Db 3818335 TGGACAACCTCATCACCGCTTCCAAAGCGCGGGAAGTGGTATCGCTGTGTACTCGA 3818276
Qy 469 laThrAspProGluAlaAsnAsnLeuValPheGlnAsnAspValGlnAspGlyTyrS 489
Db 3818275 GC-----AAGCAGGGACACG 3818261
Qy 489 erMetProGlnProThrPheArgTyrArgProSerThrAlaSerAsnValArgAlaArgL 509
Db 3818260 GCGAGCCGAACCTCATATGG-----ATCCCGATCGGCAACCGTCAACCGCTC 3818213
Qy 509 ysmMetAlaAspMetCysGluValAlaSerAsnLeuGlyGlyTyrLeuProThrSerP 529
Db 3818212 GCATCGCGCGCAAAATCGACGGCTGCGCGCGCACCTGGGTGAGTGTTCACAPCC 3818153
Qy 529 roProGlnPheMetAspProGlyLeuAlaLeuHisLeuAlaGlyThrThrArgIleGlyP 549
Db 3818152 CG-----CTCACCGCGCATTTCTCGCGCGCGCGGTGATGTCGCG 3818114
Qy 549 heAspLysAlaThrThrValAlaAspAsnAsnSerLeuValTrpAspPheAlaAsnLeu 569
Db 3818113 ACGACCCGCGACGCGGTATCATCCCTATCACCGGTCTATGGCTATCCGACGCTGT 3818054
Qy 569 yrValAlaGlyAsnGlyThrIleArgThrGlyPheGlyGluAsnProThrLeuThrSerM 589
Db 3818053 ACGTGGTTGACGGTGGCGGATCTCGGCGAACCTGGGTGTCAACCGCTGCTGTCCATCG 3817994
Qy 589 etCysHisAlaIleLysSerAla-ArgSerIleIleAsnThrLeuLysGlyGlyThrAsp 608
Db 3817993 CCGCTCAAGCGCGCGCGCGCTCGCTGT-----GGCCGGAACAAGG 3817952
Qy 609 GlyLysAsnThrGlyGluHisArgAsnLeu 618
Db 3817951 CGGAACCGCGCGCGCGCGCGCGGTG 3817922

RESULT 7

US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00


```
Db 3827288 ACGACCCCGACGCGCTCATCGACCCCTATCATCCGGGTCTATGGCTATCCGACGCTGT 3827229
Qy 569 yrValAlaGlyAsnGlyThrIleArgThrGlyPheGlyGluAsnProHisGlnArgGluPheGluA 154
Db 3827228 ACGTGGTTCAGCGTGGCGGATCTCGGGGAACCTGGGTGTCAACCCGTCGCTGCCATCG 3827169
Qy 589 etCysHisAlaIleLysSerAla-ArgSerIleIleAsnThrLeuLysGlyGlyThrAsp 608
Db 3827168 CCGCTCAAGCGGAGCGGGCCGCTCGCTGT-----GGCGGAACAAGG 3827127
Qy 609 GlyLysAsnThrGlyGluHisArgAsnLeu 618
Db 3827126 GCGAAACCGACCGCGCGCCACCGCAGGCTG 3827097

RESULT 8
US-09-252-991A-9081/c
; Sequence 9081, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 9081
; LENGTH: 1944
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (4)
; OTHER INFORMATION: Identity of nucleotide at the above locations are unknown.
US-09-252-991A-9081

Alignment Scores:
Pred. No.: 8.91e-06 Length: 1944
Score: 146.00 Matches: 147
Percent Similarity: 30.98% Conservative: 68
Best Local Similarity: 21.18% Mismatches: 271
Query Match: 4.45% Indels: 210
DB: 4 Gaps: 24

US-09-856-327-2 (1-618) x US-09-252-991A-9081 (1-1944)
Qy 15 SerMetGlnIleAsnGlyGlnIleProLysAsnAlaIleHisGluThrTyrGlyAsnAsp 34
Db 1913 TCCCTTCCCGCGGTGGCGATTCGCGGAGAGGGCTGAACATGGCGACGACACTCAA 1854
Qy 35 GlyValAspValPheIleAlaGly-SerGlyProIleGlyAlaThrTyrAlaLysLeuCy 54
Db 1853 CCGGTGGACGCGGTGGTGTCTTCGCGTGGACCGGAGCGATCCTGGCCAAAGGAGCT 1794
Qy 54 sValGluAla-GlyLeuArgValValMetValGluIleGlyAlaAlaAspSerPheTyrA 74
Db 1793 GACCGAGCGCGGGCTCAACGTGTGGCGCTGGAGCGCGGAGAACCCGATACCTATC 1734
Qy 74 LaValAsnAlaGluGlyThrAlaValProTyrValProGlyTyrHisLysLysAsnG 94
Db 1733 CCGAGCGCGCTACCCCAACACCTC-----GAGC 1704
Qy 94 luIleGluPheGlnLysAspIleAspArgPheValAsnValIleLysGlyAlaLeuGlnG 114
Db 1703 AGCTGACCTACAACACCGCGCAAGCTGTTCCAGAATCTCTCCAAGACCGCTG---- 1648
Qy 114 lnValSerValProValArgAsnGlnAsnValProThrLeuAspProGlyAlaTrpSerA 134
```

```
Db 1648 -----
Qy 134 laProProGlySerSerAlaIleSerAsnGlyLysAsnProHisGlnArgGluPheGluA 154
Db 1647 -----AGCATCCGACGCGGATCAACGACACCGCGCTGCTTACCGCC 1605
Qy 154 snLeuSerAlaGluAlaValThrArgGlyValGlyGlyMetSerThrHisThrPheCys 174
Db 1604 AGCTCTCGCGCTTCTTCCCGCGACGCGTGGTGGCGCGCTGCTGCTGCTGCTGCTGCG 1545
Qy 174 erThrProArgIleHisProMetGluSerLeu-ProGlyIleGlyArg----- 190
Db 1544 TGCATTTCCGGATCATGCCGAGGAACCTGCCGACGCAATTAACGAAACGCTACG 1485
Qy 191 -----ProLysLeu-----SerAsnAspProAlaGluAspLysGlu 203
Db 1484 CCAAGAAATTCATCCCGAAGGCATGACCATCCAGGACTACGGGTGCTACGAGAGC 1425
Qy 204 TrpAsnGluLeuTyrSerGluAlaGluArgLeuGlyThrSerThrLysGluPheAsp 223
Db 1424 TGGAGCGCAC-TTCGATTTCCCGGAGAAGGTCTTCGCGCACTCC----- 1381
Qy 224 GluSerIleArgHisThrLeu----- 230
Db 1380 ---GGCACCGCGCATACGGTCAAGGGCCAGGTGTGCGCAAGGCAATCCGTTCGCCGCC 1324
Qy 231 -----ValLeuArgSerLeuGlnAspAlaTyrLysAspArgGlnArg 244
Db 1323 GACCGCTCGGACGACTTCCCGCTGCCGCGCATGCGCAGGTGTACTCC-----CGC 1270
Qy 245 IlePheArg-----ProLeuProLeuAla--- 252
Db 1269 CTGTTCCGAAGCGCGGAGGAACCTCGGCTGCAACCTTACGACCTCGCGGCGCAAC 1210
Qy 253 -----CysHis 254
Db 1209 GCTTCGGGGCCCTGGACAAACCCCTACGGGTGCGAGATGGGGCCGTGCAACTTCGCGGC 1150
Qy 255 ArgLeuLysAsnAlaProGluTyrValGluTrpHisSerAlaGluAsnLeuPheHisSer 274
Db 1149 TTCGACGCGCTACGCGCTGTACATGTATTCGAAGGCTCGCCGAACCTG-----AAC 1096
Qy 275 IleTyrAsnAspAspLysGlnLysLysLeuPheThrLeuLeuThrAsnHisArgCysThr 294
Db 1095 ATCCTCCGCGCGCTGCGCGACACCGCGCTGTTCGAACCTGCGCGCACTGCAACGTGCTC 1036
Qy 295 ArgLeuAlaLeuThrGlyGlyTyrGluLysLysIleGlyAla-----AlaGlu 310
Db 1035 AAGGTCAACCTCGATAGCGATGGCGAGCGCCCGCGGTGACCTACGTGACGCCCGAG 976
Qy 311 ValArgAsnLeuLeuAlaThrArgAsnProSerSerGlnLeuAspSerTyrIleMetAla 330
Db 975 GGGCGCGAGATCGTGACCG----- 952
Qy 331 LysValTyrValLeuAlaSerGlyAlaIleGlyAsnProGlnIleLeuTyrAsnSerGly 350
Db 951 AAGCTGGTATCATCAGTGCCTCCAGTTCACAACGTGCGCTGCTGCTGCTGCTGCTGCTG 892
Qy 351 PheSerGlyLeuGlnValThrProArgAsnAspSerLeuIleProAsnLeuGlyArgTyr 370
Db 891 ATC---GGCAACCGCTACGACCCCGCGCGAGGCGGTG-----GTGGGGAAGAAC 841
Qy 371 IleThrGluGlnProMetAlaPheCysGlnIleValLeuArgGlnGluPheValAspSer 390
Db 840 TTCGCTTACCAAGACATGCGC-----ACCATCAAGGCGGTTTTTCGACAAGGAC 793
Qy 391 ValArgAspAspProTyr----- 396
Db 792 GTGATACCAACCCCTTTCGCGACCGCGCGCGGTGGCGGTGGAGACTTCAAC 733
Qy 397 -----GlyLeuProTyrTrpLysGlu 403
Db 732 GCCGACAACCTTCGACCGGACCGCTGGGCTTCGTGCGCGCGCTCGCCGATGTGG----- 679
```

QY 404 AlaValAlaGlnHisIleAlaLysAsnProThrAspAlaLeuProIleProPheArgAsp 423
Db 678 -----GTCAACAGCGCGCTCCAAAGCGGATCGCGCGCTCGCGCGCGCGCGCGCG 625
QY 424 ProGluProGlnValThrProPheThrGluGluHisProThrHisThr----- 440
Db 624 CCGAGC-----TGGGCGAGCGGCTGGA 601
QY 441 GlnIleHisArgAspAlaPheSerTyrGlyAlaValGlyProGluValAspSerArgVal 460
Db 600 CAGCGGTCAAGGACCTACGCGC----- 574
QY 461 IleValAspLeuArgTrpPheGly-AlaThrAspProGluAlaAsnAsnLeuValPh 480
Db 573 ACCGTGTCTGATGGAGCGCCACGCGCAGCAACATGACCTACCGCGACAAATAG-TCGACAT 515
QY 480 eGlnAsnAspValGlnAspGlyTyrSerMetPro-----GlnProThrPheArgTyr-- 497
Db 514 CGACCCGACCTACAAAGGACGCTTACGGCCAGCCACTGTTGGCGCATGACCTTCGACTGGA 455
QY 498 -----ArgProSerThrAlaSerAsnValArgAlaArgLysMetMetAlaAs 513
Db 454 CGAACACGAGATCCGATCAGCGCTACGTCACCGAGCAGCATGCGCAAGATCGCGCAAGC 395
QY 513 pMetCysGluValAla-----SerAsnLeuGlyGlyTyrLeuProThrSe 528
Db 394 GATGAACCGGAGGCGCTCGGTGAGCGTCAAGAACTTCGGCGGACCACTCAATACAGC 335
QY 528 rProGlnPheMetAspProGlyLeuAlaLeuHisLeuAlaGlyThrArgIleG1 548
Db 334 GGTCTACCAAG-----ACCACCCACCTGTCTCGCGGGGCGCATCATGGG 293
QY 548 yPheAspLysAlaThrValAlaAspAsnAsnSerLeuValTrpAspPheAlaAsnLe 568
Db 292 CAGCGATCCGAGACCGCGCTCAACCGCTACCTGACGAGCTGGGACGTGCACAACGT 233
QY 568 uTyrValAlaGlyAsnGlyThrIleArgThrGlyPheGlyGluAsnProThrLeuThrSe 588
Db 232 CTTCGTATGGCGCGCTCGGCTTCCCGCAGGCGCACCGCTACACCCCACTGGGCTGGT 173
QY 588 rMetCysHisAlaIleLysSerAlaArgSerile 599
Db 172 GCGCGCGTGGCGCTACTGCTCGGCGCAAGCGATC 139

RESULT 9

US-08-311-731A-131/C
; Sequence 131, Application US/08311731A
; Patent No. 6583266
; GENERAL INFORMATION:
; APPLICANT: SMITH, DOUGLAS
; APPLICANT: MAO, JEN-I
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; TITLE OF INVENTION: RELATING TO MYCOBACTERIUM TUBERCULOSIS AND LAPRAE FOR
; NUMBER OF SEQUENCES: 411
; CORRESPONDENCE ADDRESS:
; ADDRESSER: WOLF, GREENFIELD & SACKS, P.C.
; STREET: 600 ATLANTIC AVENUE
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: USA
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/311,731A
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:

; NAME: GATES, EDWARD R.
; REGISTRATION NUMBER: 31,616
; REFERENCE/DOCKET NUMBER: C0044/7125
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/720-3500
; TELEFAX: 617/720-2441
; INFORMATION FOR SEQ ID NO: 131:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 42325 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: MYCOBACTERIUM LEPRAE
US-08-311-731A-131
Alignment Scores:
Pred. No.: 0.125 Length: 42325
Score: 128.50 Matches: 133
Percent Similarity: 31.90% Conservative: 82
Best Local Similarity: 19.73% Mismatches: 222
Query Match: 3.91% Indels: 238
DB: 4 Gaps: 28
US-09-856-327-2 (1-618) x US-08-311-731A-131 (1-42325)

QY 22 IleProLysAsnAlaIleHisGluThrTyrGlyAsnAspGlyValAspValPheIleAla 41
Db 30466 GTCCCTAAG---GCGGGTCATACCTGGGCGCATGAAGCGGGATTATGACGTCTTAATCATC 30410
QY 42 GlySerGlyProIleGlyAlaThrTyrAlaLysLeuCysValGluAlaGlyLeuArgVal 61
Db 30409 GGTTCAGGGTTGGCGGCGATGTCAGTGGCTGCGGCTCACGAAAAAGGTTACCGGGTTC 30350
QY 62 ValMetValGluIleGlyAla-----AlaAspSerPheTyrAlaValAsnAlaGluGlu 79
Db 30349 GGTGTTTGGAGCGGCGCGCTCGCTCCGCGACGAGGATTTTGCCAGACGCTCTGGGAT 30290
QY 80 GlyThrAlaValProTyrValPro-----GlyTyrHisLysLysAsnGluIleGluPhe 97
Db 30289 CTGCGTAAATTCCTCTGGCGCGCGAAGCTGGGCTGCTACGGCATCAACGCAATTCACCTG 30230
QY 98 GlnLysAspIleAspArgPheValAsnValIleLysGlyAlaLeuGlnGlnValSerVal 117
Db 30229 CTGCGCAAC-----GTGATGATATTGGCCGCGGCC----- 30200
QY 118 ProValArgAsnGlnAsnValProThrLeuAspProGlyAlaTrpSerAlaProGly 137
Db 30200 ----- 30200
QY 138 SerSerAlaIleSerAsnGlyLysAsnProHisGlnArgGluPheGluAsnLeuSerAla 157
Db 30200 ----- 30200
QY 158 GluAlaValThrArgGlyValGlyMetSerThrHisThrThrCysSerThrProArg 177
Db 30199 -----GGAGTGGGAGGCGGTGCTGTAACACTCGCAATACG----- 30164
QY 178 IleHisProMetGluSerLeuProGlyIleGlyArgProLysLeuSerAsnAspPro 197
Db 30163 TTGTACGTTCCCGCGAG-----CCGTTCTTTGCCAACACGAGCAG 30125
QY 198 AlaGluAspAspLysGluTrp---AsnGluLeu-----TyrSerGluAlaGluArg 213
Db 30124 TGGGCGCACATCACCGACTGGCAGCAGCGCTGGCGCCGCTACGACACGCGCGCGCGG 30065
QY 214 LeuIleGlyThr-----SerThrLysGluPheAspGluSer 225
Db 30064 ATGTAGGTGTGGTTTGTAAACCGACCTTCACCGACCGCGCGGATCTCAAGAGGTT 30005


```
Db 336 CCTGATCAACTGGACC-AGCAGCTTCCGCCACCCCGGAA-----CCGACCCTGC 382
Qy 156 erAlaGluAlaValThrArgGlyValGlyGly-----MetSerThrHist 171
Db 383 AGCACTGGCCCGCCAGGTCTACGGGTCAAGGCCACAGCGCGGAGACATGGCGCCCTGGT 442
Qy 171 rpThrCysSerThrProArgIleHisProMetGluSerLeuProGlyIleGlyArgp 191
Db 443 TCGAGAAATGGAGCAGCGCTGCACGTGGCGCCCTGGCGCTGCCG-----C 490
Qy 191 roLysLeuSerAsnAspProAlaGluAspAspLysGluTrpAsnGluLeuTyrSerGluA 211
Db 491 CGAATGCCAACACGAC-----507
Qy 211 laGluArgLeuIleGlyThrSerThrLysGluPheAspGluSerIleArgHisThrLeuV 231
Db 508 -----G 508
Qy 231 alLeuArgSerLeuGlnAspAlaTyrLysAspArgGlnArgIlePheArgProLeuProL 251
Db 509 TGATCCGC-----C 517
Qy 251 euAlaCysHisArgLeuLysAsnAlaProGluTyrValGluTrpHisSerAlaGluAsnL 271
Db 518 TGGCTGCGAGAGCTCGC-----TATCACTGGAAGGTCAATCCCGCGCA 562
Qy 271 euPheHisSerIleTyrAsn-----AspAspLysG 281
Db 563 ACGTCCCTCGGTGCTGGAACTCGCTACTGCGGCATGGGCTGCCCGTCAACGCCAAGC 622
Qy 281 lnLysLysLeuPheThr-----LeuLeuThrA 290
Db 623 AGTCGATGCTGTCACCAACATCCCGCGACCCCTCGACAAAGGTGGCGAGTCTCTACC 682
Qy 290 snHisArgCysThrArgLeuAlaLeuThrGly-----G 301
Db 683 TAGCGCGCGCAACCGCTGCTGTCGACGCGCACAAAGGTACGGGCTCGAATGCTCGT 742
Qy 301 lyTyrGluLysLysIleGlyAlaAlaGluValArgAsnLeuLeuAlaThrArgAsnPro 321
Db 743 GCATGGATGACGCTGGTGGCGCCCAACGCGCCCGGATC-----783
Qy 321 erSerGlnLeuAspSerTyrIleMetAlaLysValTyrValLeuAlaSerGlyAlaIleG 341
Db 784 -----AGGTCGCGCGCCAGGCACTACGTCTCTCCGCGCGCGCATCA 826
Qy 341 lyAsnProGlnIleLeuTyrAsnSerGlyPheSerGlyLeuGlnValThrProArgAsnA 361
Db 827 ACACCCGCGCATTTCTTCGCTCGAAGCGCCTGATCC-GAGCCACGCGGTGGCAAG 885
Qy 361 spSerLeuIleProAsnLeuGlyArgTyrIleThrGluGlnProMetAlaPheCysGlnI 381
Db 886 CGCACCTT---CCTGCACAGGTGAATTTCAGCGGGGCTGTTCGACCGGGTATCAAC 942
Qy 381 leValLeuArgGlnGluPheValAspSerValArgAspProTyrGlyLeuProTrpT 401
Db 943 CCCTTCTACGCGCACCGCAGTCGATCTACTCCGACCATTTCCCAATGGAGCGGCTCGT 1002
Qy 401 rpLysGluAlaValAlaGln-----407
Db 1003 ACCGGACGCATCTCTCAAACTGGAGGTGGCGCGCTGCAACGCTGCTGGCTCGGTG 1062
Qy 408 -----HisIleAlaLysAsnProThr-----414
Db 1063 CTGCTCGCGGCTTCGGCAGGACACGCCCTGCGCATGGACAGCTACCCGATACCAAC 1122
Qy 415 --AspAlaLeuProIleProPheArgAspProGluProGln-----ValThrThrP 431
Db 1123 ATGATGTCGCGCTGCTCGCGACGGCTTCACCCCGACAGCGGAGGGCTCGGTCCGC 1182
Qy 431 ropPheThrGluGluHisProThrHisThrGlnIleHisArgAspAlaPheSerTyrGlyA 451
Db 1183 CTGCGCGCGCAGCAGACCCC---AGTGTGCTGACTACCGGATGACCCGACTACACCTGGGAC 1239
```

```
Qy 451 laValGlyProGluValAspSerArgValIleValAspLeuArgTrpPheGlyAlaThrA 471
Db 1240 GGTATCCGCGCTGCTTCCA-----1259
Qy 471 spProGluAlaAsnAsnLeuValPheGlnAsnAspValGlnAspGlyTyrSerMetP 491
Db 1260 -----TACCATGGCG 1269
Qy 491 roGlnProThrPheArgTyrArgProSerThrAlaSerAsnValArgAlaArgLysMetM 511
Db 1270 GAGATCCAGTTCGCGCGCGCGGAGAGTGGCGGCGGACAGCGCGCGGTAGC 1329
Qy 511 etAlaAspMetCysGluValAlaSerAsn-----LeuGlyGlyTyrLeuProT 527
Db 1330 CGGAAGAACCTGCGAGGAGCGCGGATCATCGACGCGCTGCGCTGGAGATCTACGCG 1389
Qy 527 hrSerProGlnPheMetAspProGlyLeuAlaLeu-HisLeuAlaGlyThrArg 546
Db 1390 ACGCGCTGGGCGAG-----CGCCACGTCATGGCGGCTGCGCC 1428
Qy 547 IleGlyPheAspLysAlaThrThrValAlaAspAsnAsnSerLeuValTrpAspPheAla 566
Db 1429 ATGGCGGAAGACCGCGCGGTGGCGGCGGACAGCTCGCGCCGCGCATCACCACCTGCGG 1488
Qy 567 AsnLeuTyrValAlaGlyAsnGlyThrIleArgThrGlyPheGlyGluAsnProThrLeu 586
Db 1489 AACCTGTCGATCCACGACGCGCTGCTGTTCCACAGCATCGGCGGCCAACCCGCGCTC 1548
Qy 587 Thr 587
Db 1549 TCG 1551
```

RESULT 11

US-09-265-108-1

; Sequence 1, Application US/09265108

; Patent No. 6033891

; GENERAL INFORMATION:-

; APPLICANT: Golightly, Elizabeth

; APPLICANT: Brown, Kimberly

; TITLE OF INVENTION: Nucleic Acids Encoding Polypeptides

; TITLE OF INVENTION: Having Cellobiose Dehydrogenase Activity

; FILE REFERENCE: 5850.000-US

; CURRENT APPLICATION NUMBER: US/09/265,108

; CURRENT FILING DATE: 1999-03-09

; NUMBER OF SEQ ID NOS: 4

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 1

; LENGTH: 2576

; TYPE: DNA

; ORGANISM: Humicola

US-09-265-108-1

Alignment Scores:

Pred. No.:	0.00366	Length:	2576
Score:	124.00	Matches:	131
Percent Similarity:	34.70%	Conservative:	89
Best Local Similarity:	20.66%	Mismatches:	270
Query Match:	3.78%	Indels:	145
DB:	3	Gaps:	30

US-09-856-327-2 (1-618) x US-09-265-108-1 (1-2576)

Qy 26 AlaIleHisGluThrTyrGlyAsnAspGlyValAspValPheIleAlaGlySerGlyPro 45

Db 794 GCCCTGAGGAGACTTAT-----GACTATATTTGTTGGTGGTGGCGCC 838

Qy 46 IleGlyAlaThrTyrAlaLysLeuCysValGluAlaGlyLeuArgValMetValGlu 65

Db 839 GCGGTATTCTGTCGCGGCAAGCTGAGCGAGCGCGCCACAGAGTTCTCTCATCGAG 898

Qy 66 IleGlyAlaAlaAspSerPheTyrAlaValAsnAlaGluGlyThrAlaValPro--- 84

Db 111


```
Db      1929  CGCGGACCTTATCGCTCCGCGCGCGC 1956
||||:|||||
RESULT 15
US-09-252-991A-11409
; Sequence 11409, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: 107196.136
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 11409
; LENGTH: 3009
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-11409

Alignment Scores:
Pred. No.: 0.00692 Length: 3009
Score: 122.50 Matches: 130
Percent Similarity: 34.95% Conservative: 65
Best Local Similarity: 23.30% Mismatches: 222
Query Match: 3.73% Indels: 142
DB: 4 Gaps: 30

US-09-856-327-2 (1-618) x US-09-252-991A-11409 (1-3009)

Qy      26 AlaIleHisGluThrTyrGlyAsnAspGlyValAspValPheIleAlaGlySerGlyPro 45
|||||
Db      365 GCGATCCACCCGCGCTACGCGC-----TTCTCTCGGAGAAC----- 400

Qy      46 IleGlyAlaThrTyrAlaLysLeuCysValGluAlaGlyLeuArgValValMetValGlu 65
|||||
Db      401 -----GCGGACTTCGCGCGCGCTGCGAGGAGCGCGCTGCTCTCTCGGACCGCG 454

Qy      66 IleGlyAlaAlaAspSerPheTyrAlaValAsnAlaGlu-----GluGlyThr 81
|||||
Db      455 GCGCGCGCATCGACCCATGCGCAGCAAGTTCGCGCGCCGCAAGCGCTGATGGAAGAGCC 514

Qy      82 AlaValProTyrValProGlyTyrHis---LysLysAsnGluIleGlu---PheGlnLys 99
|||||
Db      515 GCGCTGCCGCTGTGTGCGCGCTATCACGCGGAGAGCCAGACCTTCGAGACTTTCGCGCG 574

Qy      100 AspIleAspArg-----PheValAsnValIleLysGlyAla-LeuGlnGlnValSerVa 117
|||||
Db      575 GAAGCGCGCGCATCGCTACCGGTGCTGCTCAAGCGCGCGCTGCGCGCGCGCGCAAG 634

Qy      117 lProValArgAsnGlnAsnVal-----ProThrLeuAspProGlyAl 131
|||||
Db      635 GGCATGAAGTGTGGAGCGCGAGCGCGAGTGGCGGAGCGCTTTCCTCCGCGCAGCGC 694

Qy      131 aTrpSerAlaProProGlySerSer-AlaIleSerAsnGlyLysAsnProHisGlnArg 151
|||||
Db      695 GAGGCCAAGCGCGCTTCGCGGACGCGCTATGCTGGTGAGAAATACCT----- 744

Qy      151 luPheGluAsnLeuSerAlaGluAlaValThrArgGlyValGlyGlyMet-----SerThrH 170
|||||
Db      745 -----GCTGAAGCGCGCGCCACGTGGAGATCCAGGTATTCGCCCGCAGCGC 787

Qy      170 IsTrpThrCysSerThrProArg-----IleHisProProMetGlu 184
|||||
Db      788 CATGCCCACTGCTACTCAACGACGCGACTGTTGATCCAGCGCGCGCCACGAGAG 847

Qy      184 erLeuProGlyIleGlyArgProLysLeuSerAsnAspProAlaGluAspLysGluT 204
|||||
||||:|||||
```

Search completed: September 24, 2003, 10:58:43
Job time : 7718 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.
OM protein - nucleic search, using frame_plus_p2n model

Run on: September 24, 2003, 05:33:29 ; Search time 3509 Seconds
(without alignments)
4280.469 Million cell updates/sec

Title: US-09-856-327-2
Perfect score: 3284
Sequence: 1 MSLSTEQMLRDPYRSMQING.....IINTLKGTGDKNGTGEHRNL 618

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n_model -DEV=xlh
-O=/cgn2_1/USPTO_spool/US09856327/runat_23092003_153419_15517/app_query.fasta_1.775
-DB=EST -QFMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-LOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09856327.ecgn.1.1.2810 -runat_23092003_153419_15517 -NCPU=6 -ICPU=3
-NO_MMMap -LARGQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_estl:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rpd:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gssl:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	241	7.3	449	9	AI213008
2	120.5	3.7	1971	13	BU227568 y6e01al.r
3	117.5	3.6	384	29	CJE315007 Campyloba
4	112.5	3.4	1145	12	BM462502 AGENCOURT
5	112	3.4	834	10	BG541504 602570950
6	112	3.4	839	28	BH451002 BCGY12TR
7	109	3.3	1451	11	AK029462 Mus muscu
8	109	3.3	2996	11	BC011088 Mus muscu
9	107	3.3	690	12	BJ258831 BJ258831
10	107	3.3	1319	13	BU518146 AGENCOURT
11	106.5	3.2	3111	11	AK085667 Mus muscu
12	106	3.2	691	9	AV764246 AV764246
13	106	3.2	783	28	BH494670 BOGV59TR
14	106	3.2	920	10	BE782505 601467649
15	106	3.2	4866	11	BC034946 Homo sapi
16	105.5	3.2	605	12	BM425539 pgf2c.pk0
17	105.5	3.2	1201	9	AL536149 AL536149
18	105.5	3.2	1288	13	BU541981 AGENCOURT
19	104	3.2	2518	11	BC003719 Mus muscu
20	103.5	3.2	976	13	BQ713154 AGENCOURT
21	103.5	3.2	988	11	CNS08JNK BX014892 Single re
22	103	3.1	503	9	AW255795 ML863 pep
23	103	3.1	661	12	BJ136654 BJ136654
24	103	3.1	1003	13	BU962960 AGENCOURT
25	103	3.1	1819	11	BC031681 Homo sapi
26	103	3.1	1869	11	BC040952 Homo sapi
27	103	3.1	1924	10	BG027879 BG027879
28	102.5	3.1	1023	13	BU124024 603147663
29	102.5	3.1	1514	11	BC051546 Mus muscu
30	102.5	3.1	2910	11	AK078478 Mus muscu
31	102	3.1	633	9	AL957104 AL957104
32	102	3.1	795	28	BH708030 BOMDQ12TF
33	102	3.1	886	28	BH972564 odh12a06
34	102	3.1	3611	11	AK053309 Mus muscu
35	101.5	3.1	766	12	BM291983 EST574525
36	101.5	3.1	1201	13	BX423637 BX423637
37	101	3.1	707	12	BM609508 170006871
38	101	3.1	733	28	BH524476 BGGOK35TF
39	101	3.1	822	10	BF500573 AT15559.5
40	101	3.1	905	10	BE616603 601278816
41	101	3.1	977	29	CC357805 PURPH38TD
42	100.5	3.1	732	28	BH601509 BGSB77TF
43	100.5	3.1	1031	12	BM469849 AGENCOURT
44	100.5	3.1	1230	29	CC187722 CC187722
45	100.5	3.1	2592	11	AK045276 Mus muscu

ALIGNMENTS

RESULT 1
AI213008
LOCUS y6e01al.r1 Aspergillus nidulans 24hr asexual developmental and vegetative cDNA lambda zap library Emericella nidulans cDNA clone y6e01al 5', mRNA sequence.
DEFINITION AI213008
ACCESSION AI213008
VERSION AI213008
KEYWORDS EST
SOURCE Emericella nidulans (anamorph: Aspergillus nidulans)
ORGANISM Emericella nidulans
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes; Eurotiales; Trichocomaceae; Emericella.

```

REFERENCE 1 (bases 1 to 449)
AUTHORS  Kupfer,D., Gray,J., Hausner,J., Lai,H., Martin,W., Aramayo,R.,
          Prade,R. and Roe,B.
TITLE     An Aspergillus nidulans EST Database
JOURNAL   Unpublished
COMMENT   Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu
          Department of Chemistry and Biochemistry
          Advanced Center for Genome Technology, University of Oklahoma
          620 Parrington Oval, Norman, OK 73019, USA
          Tel: 405 325 4912
          Fax: 405 325 7762
          Email: broe@ou.edu
          We anticipate the future release of the cDNA clones to the Fungal
          Genetics Stock Center
          Seq primer: T3
          High quality sequence stop: 442.
          Location/Qualifiers
            source          1..449
                           /organism="Emmericella nidulans"
                           /mol_type="mRNA"
                           /strain="FGSC A26"
                           /db_xref="taxon:162425"
                           /clone="y6e01a1"
                           /tissue_type="vegetative mycelia, asexual structures"
                           /clone_lib="Aspergillus nidulans 24hr asexual
                           developmental and vegetative cDNA lambda zap library"
                           /notes="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                           XhoI; 5' end of cDNA cloned into EcoRI site of pBluescript
                           3' end of cDNA cloned into XhoI site of pBluescript"
BASE COUNT. 130 a 108 c 115 g 96 t
ORIGIN
Alignment Scores:
Pred. No.:      3.05e-13      Length:      449
Score:          241.00       Matches:     58
Percent Similarity: 52.90%    Conservative: 15
Best Local Similarity: 42.03% Mismatches:    41
Query Match:     7.34%      Indels:      24
DB:              9          Gaps:         3

US-09-856-327-2 (1-618) x AI213008 (1-449)
Qy  37 AspValPheIleAlaGlySerGlyProIleGlyAlaThrTyrAlaLysLeuCysValGlu 56
      ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  80 GATGTCCTAATCATCGGTCGGCCCGCCAGTCGGGGCAACTTACCGAGGGAGATTCTGGAC 139
Qy  57 AlaGlyLeu-----ArgValValMetValGluIleGly 67
      |||  |||  ::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db  140 CCCGGCTCCGGTGCTTCGCCTGGAGAAAAGCTCCAAAGGTTATTATGTGGAAACTGGG 199
Qy  68 AlaAlaAspSerPheTyrAlaValAsnAlaGluGlyThrAlaValProTyrValPro 87
      |||  ::|||
Db  200 GCCACGAGGTC-----AAAGTCCT 220
Qy  88 GlyTyrHisLysLysAsnGluIleGluPheGlnLysAspIleAspArgPheValAsnVal 107
      ||||| ||||| ||||| ::| ::| ||||| ||||| ||||| ||||| |||||
Db  221 GGTGAACACAGAAGAAATGCCGTGGTATACCAACAACACATGATCTCTTTGCAATGTT 280
Qy  108 IleLysGlyAlaLeuGlnValSerValProValIargAsnGlnAsnValProThrLeu 127
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  281 ATTCAGGGAGCGCTATTTCGAACCTCTGTCCCAACACAGCTGTGGATCCGAACCTCAAGTTG 340
Qy  128 AspProGlyAlaTrpSerAlaProProGlySerSerAlaIleSerAsnGlyLysAsnPro 147
      |||  ::| ||||| |||||
Db  341 CTTCCCGTCCTCTGGTCT-----CCACGTGAAAAACAAACCTTCAACGGCGCAACAAG 394
Qy  148 HisGlnArgGluPheGluAsnLeuSerAlaGluAlaValThrArgGlyValGly 165
      |||  ::| ||||| |||||
Db  395 GACCAGAATATCTACCACAACTTAGCGCTAATGGTGTGTCCCGCAATGTTGGC 448

RESULT 2
LOCUS    BU227568
LINEAR   1971 bp      mRNA      linear      EST 26-NOV-2002

```

```

DEFINITION 603947501f1 CSEQCHN23 Gallus gallus cDNA clone CHEST902b2 5', mRNA
sequence.
ACCESSION BU227568
VERSION BU227568.1 GI:25465582
KEYWORDS EST.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
1 (bases 1 to 1971)
Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E.,
Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
A Comprehensive Collection of Chicken cDNAs
Curr. Biol. 12 (22), 1965-1969 (2002)
22335534
12445392
Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology (UMIST
)
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.
Location/Qualifiers
1. 1971
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="White Leghorn, Hisex"
/db_xref="taxon:9031"
/clone="CHEST902b2"
/dev_stage="22"
/lab_host="DH10B"
/clone_lib="CSEQCHN23"
/notes="Organ: heads; Vector: pBluescript II KS(+); Site:1:
ECORI; Site 2: NotI; This normalized library was
constructed from 1 million independent clones. cDNA
synthesis was initiated using an oligo(dT) primer, using
methylated C in the first strand synthesis reaction.
Following this first strand reaction, double-stranded cDNA
was blunted, ligated to NotI adapters, digested with ECORI
, size-selected, and cloned into the NotI and ECORI
compatible sites of a custom modified MCS of the
pBluescript (KS-) vector. The library was normalized in 2
rounds using conditions adapted from Soares et al., PNAS
(1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6
(1996): 791, except that a significantly longer
reannealing hybridization was used."
BASE COUNT 564 a 847 c 370 g 190 t
ORIGIN

Alignment Scores:
Pred. No.: 1.04 Length: 1971
Score: 120.50 Matches: 79
Percent Similarity: 33.74% Conservative: 59
Best Local Similarity: 19.32% Mismatches: 180
Query Match: 3.67% Indels: 91
DB: 13 Gaps: 12

US-09-856-327-2 (1-618) x BU227568 (1-1971)

Qy 116 ServValProValArgAsnGlnAsnValProThrLeuAspProGlyAlaTrpSerAlaPro 135
::: ||| ::: ||||| ||||| ::: |||
755 AACACCCCAACACCGGCCCAACCGCCCAACTCGCGACCTGGCGCCCAACGACGCCCC 814
::: |||

Qy 136 ProGlySerSerAlaIleSerAsnGlyLysAsnProHisGlnArgGluPheGluAsnLeu 155
||| ::|||::: ::|
Db 815 CAGGGCAGACAGCGCCCAACACCGGCCCAACCCCTACACCCACGCGCGCACACAC 874
::|

Qy 156 SerAlaGluAlaValThrArgGlyVal-----GlyGlyMetSerThrHisTrpThrCys 173
::: ||| |||||::: ::|
875 AACCCCCCAACAAACAAAGACCTTCGGCGCAACCTTCGGCGCAACACGCGCTCTCT 934
::: ||| |||||::: ::|

```

DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
COMMENT
FEATURES
SOURCE
BASE COUNT
ORIGIN
Alignment :
Pred. No. :
Score:
Percent Sim
Best Local
Query Match
DB:
US-09-856-
Qy
Dd
Qy
Dd
Qy
Dd

Fax: 301-838-0208		Email: cdtown@tigr.org	
DNA is from a doubled haploid provided by Tom Osborn.			
Seq primer: TR		Class: sheared ends.	
FEATURES		Location/Qualifiers	
source			
1..839			
/organism="Brassica oleracea"			
/mol_type="genomic DNA"			
/strain="Tol1000DH3"			
/db_xref="taxon:3712"			
/clone="BOGXY12"			
/clone_lib="BOGX"			
/note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared genomic DNA inserted into pHOS1 using BstXI linkers"			
BASE COUNT	250 a	136 c	221 g 232 t
ORIGIN			
Alignment Scores:			
Pred. No.:	2.14	Length:	839
Score:	112.00	Matches:	59
Percent Similarity:	32.10%	Conservative:	28
Best Local Similarity:	21.7%	Mismatches:	78
Query Match:	3.41%	Indels:	106
DB:	28	Gaps:	12
US-09-856-327-2 (1-618) x BH451002 (1-839)			
QY	183	GlUSeRLeuProGlyIleGlyArgProLysLeuSerAsnAspProAlaGluAsp	---Asp 201
Db	8	GAATGCTTCTCGGGCGGAATAACCTCCATATGGAAGACAAAATTCAGACCTGCAG	67
QY	202	LysGluTrpAsnGluLeuTyrSerGluAlaGluArgLeuIleGlyThrSerThrLysGlu	221
Db	68	AGACATTCGAGAGGTCCATCCGCACCAACATAGG	115
QY	222	PheAspGluSerIleArgHisThrLeuValLeuArgSerLeuGlnAspAlaTyrLysAsp	241
Db	116	GTTTGTAGG	157
QY	242	ArgGlnArgIlePheArgProLeuProLeuAlaCysHisArgLeuLysAsnAlaProGlu	261
Db	158	GAGGAG	166
QY	262	TyrValGluTrpHis	280
Db	167	TAC	211
QY	281	GlnLysLysLeuPheThrLeuLeuThrAsnHisArgCysThrArgLeuAlaLeuThrGly	300
Db	212	CTAACAAATTTTATCATGCCTTGACAAAGCAAGACGGGTTCGGAATAGGATTGTGGGT	271
QY	301	GlyTyrGlu	308
Db	272	CTTCATGATGAGCGGGTACTTGGATAACAGGACGAAGGTGTGGAAGAGTGGCTGTG	331
QY	309	AlaGluValArgAsnLeuAlaThrArgAsnProSerSerGlnLeuAspSerTyrIle	328
Db	332	GAATATTTGAGGATTTATTTAGTACCACTTCTCCA	367
QY	329	MetAlaLysValTyrValLeuAlaSerGlyAlaIleGlyAsnProGlnIleLeuTyrAsn	348
Db	367		367
QY	349	SerGlyPheSerGlyLeuGlnValThrProArgAsnAspSerLeuIleProAsnLeuGly	368
Db	368	TCAGGTTTT	400
QY	369	ArgTyrIleThrGluGlnProMetAlaPheCysGlnIleValLeuArgGlnGluPheVal	388
Db	401	ATGCTTATTACTCTCCTCAG	451
QY	389	AspSerValArgAspProTyrGlyLeuProTyrIleValGluAlaValAlaGlnHis	408

```

452 Db      452 GAGGAAGTTTCGTCAG      456
      :::: |||||
409 Qy      409 IleAlaLysAsnProThrAspAlaLeuProIleProPheArgAspProGluProGlnVal 438
      |||| :::: |||| :::: ||||
467 Db      467 -----ACTTTGTATGATGATCGGNAAGATCGCGGACCGGATGG 511
      :::: |||| :::: ||||
429 Qy      429 ThrThrProPheThrGluGluHisProTrpHis 439
      ||| :::: |||||
512 Db      512 ATGACAGCTCTTTTTCACACACTCATGGCAT 544
      ||| :::: |||||

RESULT 7
AK029462
LOCUS      AK029462
DEFINITION Mus musculus 0 day neonate head cDNA, RIKEN full-length enriched
            library, clone:483344J24 product:HOTTL PROTEIN homolog [Homo
            sapiens], full insert sequence.
ACCESSION AK029462
VERSION    1451 bp mRNA linear HTC 05-DEC-2000
KEYWORDS   Mus musculus 0 day neonate head cDNA, RIKEN full-length enriched
SOURCE     library, clone:483344J24 product:HOTTL PROTEIN homolog [Homo
            sapiens], full insert sequence.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus; Mus.
1
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
99279253
10349636
2
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new gene
Genome Res. 10 (10), 1617-1630 (2000)
20499374
11042159
3
Shibata, K., Itoh, M., Aizawa, K., Nagasaka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (KISA) system--384-format
Sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
11076861
4
Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,
Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S.,
Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I.,
Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R.,
Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,
Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H.,
Kuehl, P., Lewis, S., Matsuo, Y., Nikaide, I., Pesole, G.,
Quackenbush, J., Schriml, L.M., Staib, F., Suzuki, R., Tomita, M.,
Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,
Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N.,
Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Bult, C.,
Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D.,
Hofmann, M., Hume, D.A., Kamiya, M., Lee, N.H., Lyons, P.,
Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P.,
Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,
Sato, K., Schonbach, C., Seva, T., Shibata, Y., Storch, K.F., Suzuki, H.,
Toyooka, K., Wang, K.H., Weitz, C., Whitaker, C., Wilming, L.,
Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohsuke, S.,
and Hayashizaki, Y.
Functional annotation of a full-length mouse cDNA collection
Nature 409 (6821), 685-690 (2001)

```

MEDLINE PUBMED REFERENCE AUTHORS

21085660
11217851

5

TITLE JOURNAL REFERENCE AUTHORS

The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II team.
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 1451)
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.

TITLE JOURNAL

Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@sc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

COMMENT

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/.

FEATURES source

1. .1451
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:4833441J24"
/db_xref="taxon:10090"
/clone="4833441J24"
/tissue_type="head"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="0 day neonate"
329. .1192
/note="unnamed protein product; HOTTL PROTEIN homolog [Homo sapiens] (SPTRIQ9UI99, evidence: FASTV, 85.8%ID, 93.4%length, match=821)
putative"
/codon_start=1
/protein_id="BAC26459.1"
/db_xref="GI:26325410"
/translation="MGLRNNAKIHVERAVKOKIFMIHGYPVIRCLLRQGWVEKMHVPPCTALPAQKDLSDSMLGNSDATEDDEENEMFRESOLLDDGFLFDDLDGIIH ALMSRMVNETPYLIWTRRDVLDRCFLSKDOMINHYARAGSFTTKVGLCLNRLNPW FDEADADFFRCVRLGAEDEKKAFIGKGTPIGLAPSCCLSHRGEKQASPCGLCNPS PFLPLMGLPFIISQPMHSEKPREQGLYPVAGSRGGMVGSWSLKLALLPGSTKNGSL RQSKFLFOHGF"
1438. .1443
/note="putative"
1451
/note="putative"

CDS

US-09-856-327-2 (1-618) x AK029462 (1-1451)
QY 178 lleHisProMetGluSerLeuProGly-----IleGlyArgProLysLeu 193
DB: :::::::::::::: :::::::::: :::::
461 GTGCATCCCCAGGCACAGCCCTGCCAGCACCCAGAGGATCTCGATAGCTCCATGCTG 520
QY 194 SerAsnAspProAla-----GluAspAspLysGluThrAsnGluLeuTyrSerGluAla 211
DB: :::::::::::::: :::::::::: :::::
521 GGTAAATACGATGCCAGCAGGATGAGGATGAGGAGAGAGAGAGAGAGAAATGTTCCGGAGTC 580
QY 212 GluArgLeuIleGlyThrSerThrLysGluPheAspGlu---SerIleArgHisThrLeu 230
DB: :::::::::::::: :::::::::: :::::
581 CAGCTGTGGAGCTTGGATGGTTCTCGGAATTTGATGACCTAGATGGGATACATGCTTTG 640
QY 231 ValLeuArgSerLeuGlnAspAlaTyrLysAspArgGlnArgIlePheArgProLeuPro 250
DB: :::::::::::::: :::::::::: :::::
641 ATGTCCCGC----- 649
QY 251 LeuAlaCysHisArgLeuLysAsnAlaProGluTyrValGluTyrHisSerAlaGluAsn 270
DB: :::::::::::::: :::::::::: :::::
650 -----ATGTTTCGAAATGAGACCCCTACCTACCTATCGACCACTGGCGGGAT 697
QY 271 LeuPheHisSerIleTyrAsnAspLysGlnLysLeuPheThrLeuLeuThrAsn 290
DB: :::::::::::::: :::::::::: :::::
698 GTGCTGGATTGCGCTTCTTCCAGGATCAG-----ATGATAAAC 739
QY 291 HisArgCysThrArgLeuAlaLeuThrGlyGlyTyrGluLysLysIleGly---AlaAla 309
DB: :::::::::::::: :::::::::: :::::
740 CAT-----TATGCCGTCGAGCTCTTCCACCAAGAGTGGCGCTTTGTCCTC 787
QY 310 GluValArgAsnLeuAlaThrArgAsnPro-----SerSerGlnLeuAspSer 326
DB: :::::::::::::: :::::::::: :::::
788 AACCTCCGGAATCTG-----CGTGTGTTGTAGAGCGCGATCGTCACTCT 832
QY 327 TyrIleMetAlaLysValTyrValLeuAlaSerGlyAlaIleGlyAsnProGlnIleLeu 346
DB: :::::::::::::: :::::::::: :::::
833 TTC---TTCCACGATGTACCGATTG-----GGGCGAGGATGACAGAAAGCCCTTC 883
QY 347 TyrAsnSerGlyPheSerGlyLeuGlnValThrProArgAsnAspSerLeuIleProAsn 366
DB: :::::::::::::: :::::::::: :::::
884 ATAGTAAGGG-----ACTCCA-----ATCCAGCG 910
QY 367 LeuGlyArgTyrIleThrGluGlnProMetAlaPheCysGlnIleValLeuArgGlnGlu 386
DB: :::::::::::::: :::::::::: :::::
911 CTGGCT-----CCCTCATGCTGTCTGCCACAGG----- 940
QY 387 PheValAspSerValArgAspAspProTyrGlyLeuProTyrTrpLysGluAlaValAla 406
DB: :::::::::::::: :::::::::: :::::
941 -----GGAAAGACCAAGCCCTCCCTGTGGCCCTC----- 970
QY 407 GlnHisIleAlaLysAsnProThrAspAlaLeuProIleProPheArgAspProGluPro 426
DB: :::::::::::::: :::::::::: :::::
971 -----TGTAACCCCGCCCTTCTCTCTACCA----- 997
QY 427 GlnValThrThrProPheThrGluGlnHisProThrHisThrGlnIleHisArgAspAla 446
DB: :::::::::::::: :::::::::: :::::
998 ---ATGGGTCTTCCATTCAAGCCAA---CCTTGGCATTCAGAGAAAGCCTAGGAGGCAA 1051
QY 447 PheSerTyr-----GlyAlaValGlyProGluValAspSerArgValIleValAsp 463
DB: :::::::::::::: :::::::::: :::::
1052 GGCTGTACCTGTAGCGGGTCTAGAGGTGGGTGGTGGTCTTCTGTCTGGAGTCTAGAT 1111
QY 464 LeuArgTTrpPheGlyAlaThrAspProGluAlaAsnAsnLeuLeu----- 478
DB: :::::::::::::: :::::::::: :::::
1112 AAGTGTCTGGTGGACCGAGGTCTACAAAGAGCATGGTCTCTAAGCAAGAAAGTAAGTTC 1171
QY 479 ValPheGlnAsn-----AspValGlnAspGlyTyr--- 488
DB: :::::::::::::: :::::::::: :::::
1172 CTGTTTCAACATGGGTTCTGAGTTTTCAGAAATTTGGATTCAGATCAGATCAGACGCTTTTAA 1231

polya_signal

1438. .1443

polya_site

1451

BASE COUNT

348 a 364 c 398 g 341 t

ORIGIN

Alignment Scores:

Pred. No.: 8.87 Length: 1451

Score: 109.00 Matches: 81

Percent Similarity: 34.92% Conservative: 44


```

Db 2723 -----AGACCCATGAGCGGTGAGCATCTGAGCCCTGGTCCCATCCCTGTTTCCAGG 2775
Qy 419 ePro-----PheArgAspProGluProGlnValThrThrProPheThrGluGluHis-- 436
Db 2776 GCCAGGACTATTTCAGTCAGTACCTTCACATGGAAACAGCCAGACTGGAGGAACCATGA 2835
Qy 437 ----ProTrp 438
Db 2836 GGGGCGCTGG 2845

RESULT 9
BJ258831
LOCUS BJ258831 690 bp mRNA linear EST 08-APR-2002
DEFINITION BJ258831 Y. Ogihara unpublished cDNA library, Wh_h Triticum
aestivum cDNA clone whh9m05 5', mRNA sequence.
ACCESSION BJ258831
VERSION BJ258831.1 GI:20081455
KEYWORDS EST.
SOURCE Triticum aestivum (bread wheat)
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
; Triticeae; Triticum.
1 (bases 1 to 690)
Ogihara, Y. and Murai, K.
Expressed genes in Triticum aestivum
Unpublished
Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
Location/Qualifiers
1..690
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="Chinese Spring"
/db_xref="taxon:4565"
/clone="whh9m05"
/tissue_type="spike at heading date"
/dev_stage="Peekes", scale 10.5"
/clone_lib="Y. Ogihara unpublished cDNA library, Wh_h"
BASE COUNT 159 a 190 c 189 g 152 t
ORIGIN

Alignment Scores:
Pred. No.: 4.99 Length: 690
Score: 107.00 Matches: 51
Percent Similarity: 40.94% Conservative: 19
Best Local Similarity: 29.82% Mismatches: 64
Query Match: 3.26% Indels: 37
DB: 12 Gaps: 9

US-09-856-327-2 (1-618) x BJ258831 (1-690)

Qy 120 ArgAsnGlnAsnValProThrLeuAspProGlyAlaTrpSerAlaProGlySerSer 139
Db 77 AGGAGGAGTGGCGTGGCG-----TCCCTCCCGGAGCAGTCC 115
Qy 140 AlaIleSerAsnGlyLysAsnProHisGlnArgGluPheGluAsnLeuSerAlaGluAla 159
Db 116 GCATCTCCCGCTCAAGGCGCACCGAGTTACGGGACAGCGGAGTACAAAGTCTACG 175
Qy 160 ValThrArgGlyValGlyGlyMetSerThrHisTrpThrCysSerThrProArgIleHis 179
Db 176 GTGACG-----GGCGGTACAACTCGCTGGTGGTGGGAAACGCCCTGTACAAAGT 223
Qy 180 ProProMetGluSerLeuProGlyIle---GlyArgPro-----LysLeuSerAsnAsp 196
Db 224 CCACCACCAAAATTCGACTCCGCTGGCTGGCGGCTTCTTCGAGGGGCTCCCTGGAG 283

```

```

Qy 197 Pro-AlaGluAspAsp-----LysGluTrpAsnGluLeuTyrSerGluAlaG1 212
Db 284 CCATACACAGGACGCCCTGACCCCTGATGGTAGGAGGTAGATCACATGCCATCTTGGC 343
Qy 212 uArgLeuIleGlySerThrLysGluPheAspGluSerIleArgHisThrLeuValLe 232
Db 344 CGGGCATTTGGGACATGTGTCAAGG----- 371
Qy 232 uArgSerLeuGlnAspAlaTrpLys-----AspArgGlnArgIlePheArgProLe 249
Db 372 -AGAAGCTTCAAGACGCCCTACCGATGAGCCCACTGGCTCAACAGCGTTTCCATCAAGT 430
Qy 249 uProLeuAlaCysHisArgLeuLysAsnAlaProGluTrpValGluTrpHisSerAlaG1 269
Db 431 TCAC---TCCTGCCCTCCTCTGTAATCCGGCCAGTTCGAGTCACTCTGG-----AT 478
Qy 269 uAsnLeuPheHisSerIleTyrAsnAspAsp 279
Db 479 GAACCTGTTTAAACAATGCAATAATGAT 509

RESULT 10
BU518146
LOCUS BU518146 1319 bp mRNA linear EST 12-SEP-2002
DEFINITION AGENCOURT_10166431 NIH_MGC_134 Mus musculus cDNA clone
IMAGE:6515841 5', mRNA sequence.
ACCESSION BU518146
VERSION BU518146.1 GI:22825672
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1319)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgaabs-r@mail.nih.gov
Tissue Procurement: Dr. David Rowe
CDNA Library Preparation: Invitrogen Corp
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM14093 Row: g Column: 10
High quality sequence start: 177
High quality sequence stop: 557.
Location/Qualifiers
1..1319
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:6515841"
/tissue_type="undifferentiated limb"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_134"
/notes="vector: pCMV-SPORT6.1.cdb; Site_1: EcoRV; Site_2:
NotI; Cloned unidirectionally. Primer: Oligo dr. Average
insert size 1.7 kb. Constructed by ResGen, Invitrogen
Corp. Note: this is a NIH_MGC Library."
BASE COUNT 277 a 400 c 331 g 302 t 9 others
ORIGIN

Alignment Scores:
Pred. No.: 12.1 Length: 1319
Score: 107.00 Matches: 105
Percent Similarity: 31.60% Conservative: 47
Best Local Similarity: 21.83% Mismatches: 174
Query Match: 3.26% Indels: 158
DB: 13 Gaps: 25

US-09-856-327-2 (1-618) x BU518146 (1-1319)

```



```
QY 504 -----AsnValargAlaArgLys-----MetMetAlaAspMetCysGluValAla 518
    |||:::|||||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||
Db 2087 GGAGCGCTGCCACAGTCCAGGCCGTAAGCCCTCATGGCAGATACACTGGCGGTCCTC 2028
    |||:::|||||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||
QY 519 SerAsnLeuGlyGlyTyrLeuProThrSerProPro-----530
    |||:::|||||:::|||:::|||:::|||:::|||:::|||:::|||:::|||
Db 2027 ACAGCGCGCTTGCCCATGGCAATCACTGGGGCACCTTTGCTCTTTCGACGTCTTGGCAGT 1968
    |||:::|||||:::|||:::|||:::|||:::|||:::|||:::|||:::|||
QY 531 ---GlnPheMetAspProGlyLeuAlaLeuHisLeuAla-----GlyThrThrArgIle 547
    |||:::|||||:::|||:::|||:::|||:::|||:::|||:::|||:::|||
Db 1967 GAAGCCCTCGTGGCAGATCGATGGCCATTCACAGCGGCCGATGCCCGTGGCAGTCACT 1908
    |||:::|||||:::|||:::|||:::|||:::|||:::|||:::|||:::|||
QY 548 GlyPheAspLysAlaThrThrValAlaAspAsnAsnSerLeuValTirPaspPheAlaAsn 567
    |||:::|||||:::|||:::|||:::|||:::|||:::|||:::|||:::|||
Db 1907 GGGCGAGGAAGACTCAGCACAGTCAAGGCCAGTGAAGCCATCCTCACAGATGCAGTGTCC 1848
    |||:::|||||:::|||:::|||:::|||:::|||:::|||:::|||:::|||
QY 568 LeuTyrValAla-----GlyAsnGlyThrIleArgThrGlyPhe 580
    |||:::|||||:::|||:::|||:::|||:::|||:::|||:::|||:::|||
Db 1847 GTCTACACAGCGCCCGCTGGCTAGCTCCCGGGGACAGCGCGCTCTCTCGACGTCTTC 1788
    |||:::|||||:::|||:::|||:::|||:::|||:::|||:::|||:::|||
QY 581 GlyGluAsnProThrLeuThrSerMetCysHisAlaIleLysSerAlaArgSerIleIle 600
    |||:::|||||:::|||:::|||:::|||:::|||:::|||:::|||:::|||
Db 1787 CCCAGTGTAGTCTCATCAGAT---ACACATGCCATTTCACACAGCGGCCATGTGGTG 1731

RESULT 12
AV764246 691 bp mRNA linear EST 19-OCT-2000
LOCUS AV764246 MDS Homo sapiens cDNA clone MDSGB05 5', mRNA sequence.
ACCESSION AV764246
VERSION AV764246.1 GI:10922094
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 691)
Gu., J., Zhao, M., Huang, Q., Xu, X., Li, Y., Peng, Y., Song, H., Xiao, H.,
Gu., Y., Li, N., Qian, B., Liu, F., Qu, J., Gao, X., Cheng, Z., Xu, Z., Zeng
L., Xu, S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu, G.,
Yang, Y., Gao, G., Zhang, Q., Chen, S., Han, Z. and Chen, Z.
Homo sapiens cDNA MDS clones
Unpublished
Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.

FEATURES
         Location/Qualifiers
             1..691
                 /organism="Homo sapiens"
                 /mol_type="mRNA"
                 /db_xref="taxon:9606"
                 /clone="MDSGB05"
                 /tissue_type="Bone marrow"
                 /cell_type="CD34+ hematopoietic stem/progenitor cell"
                 /lab_host="PM25.8"
                 /clone_lib="MDS"
                 /note="Vector: pTriplex2; Site_1: sflIA; Site_2: sflIB"

BASE COUNT 178 a 176 c 173 g 164 t
ORIGIN

Alignment Scores:
Pred. No.: 6.25 Length: 691
Score: 106.00 Matches: 69
Percent Similarity: 37.97% Conservative: 51
Best Local Similarity: 21.84% Mismatches: 94
Query Match: 3.23% Indels: 103
Db: 9 Gaps: 15
```

```
US-09-856-327-2 (1-618) x AV764246 (1-691)
QY 152 PheGluAsnLeuSerAla-----GluAlaValThrArgGlyValGlyGlyMetSerThr 169
    |||:::|||||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||
Db 15 TTTAATCACTCTGCAGCTTTGGAGAAGCAGGTTCAAGAGCTGCAGCGGGGCCACTAAA 74
    |||:::|||||:::|||:::|||:::|||:::|||:::|||:::|||:::|||
QY 170 HisTrpThrCysSerThrProArgIleHisProMetGluSerLeuProGlyIleGly 189
    |||:::|||||:::|||:::|||:::|||:::|||:::|||:::|||:::|||
Db 75 TATAGATGTGTGCTCC-----AGGAGCTCCCGCAGGAATGGA 110
    |||:::|||||:::|||:::|||:::|||:::|||:::|||:::|||:::|||
QY 190 ArgProLysLeuSerSerAsnAspProAlaGluAspAspLysGluTrpAsnGluLeuTyrSer 209
    |||:::|||||:::|||:::|||:::|||:::|||:::|||:::|||:::|||
Db 111 GTTCACAGAAAGAGT---GACCT---GAGGAAGTAAAGGAACCCAGCAAGCTTTTCT 164
    |||:::|||||:::|||:::|||:::|||:::|||:::|||:::|||:::|||
QY 210 GluAlaGluArgLeuIleGlyThrSerThrLysGluPheAspGluSerIleArgHisThr 229
    |||:::|||||:::|||:::|||:::|||:::|||:::|||:::|||:::|||
Db 165 GAAGCTCAGCACGACTATGCAACACAGAGAGGAAGTGAATGAA-----209
    |||:::|||||:::|||:::|||:::|||:::|||:::|||:::|||:::|||
QY 230 LeuValLeuArgSerLeuGlnAspAlaTyrLysAspArgGlnArgIlePhe-----Arg 247
    |||:::|||||:::|||:::|||:::|||:::|||:::|||:::|||:::|||
Db 210 -----TTAAGGAAGCTGCTGGAAGAAGAACAGAGAC---CAAAGAGTGGCTGCTGAGAAT 260
    |||:::|||||:::|||:::|||:::|||:::|||:::|||:::|||:::|||
QY 248 ProLeuProLeuAlaCysHisArgLeuLysAsnAlaProGluTyrValGluTrpHisSer 267
    |||:::|||||:::|||:::|||:::|||:::|||:::|||:::|||:::|||
Db 261 GCTCTCTGTGGCGAGAGCAGATCAGACGGTTA---GAGCAGAGTGAATGGGACTCT 317
    |||:::|||||:::|||:::|||:::|||:::|||:::|||:::|||:::|||
QY 268 AlaGluAsnLeuPheHisSerIleTyrAsnAspAspLysGlnLysLysLeuPheThrLeu 287
    |||:::|||||:::|||:::|||:::|||:::|||:::|||:::|||:::|||
Db 318 TCCCGGACTCCTATCATTTGGCTCCTGTGGCACTCAGGAGCAGGCACCTGTTATAGAT--- 374
    |||:::|||||:::|||:::|||:::|||:::|||:::|||:::|||:::|||
QY 288 LeuThrAsnHisArgCysThrArgLeuAlaLeuThrGlyGlyTyrGluLysLysIleGly 307
    |||:::|||||:::|||:::|||:::|||:::|||:::|||:::|||:::|||
Db 375 CTTACAAGCAACAGTTGTCGAAGGACCCGGAGTGGCTGGATGGAAGCGAGTCTGTG--- 431
    |||:::|||||:::|||:::|||:::|||:::|||:::|||:::|||:::|||
QY 308 AlaAlaGluValArgAsnLeuLeuAlaThrArgAsnProSerSerGlnLeuAspSerTyr 327
    |||:::|||||:::|||:::|||:::|||:::|||:::|||:::|||:::|||
Db 432 -----CGTTCACTTCTGCTTCATTCACGGAC---CCGAGTGCCCATCTTACGAGCCAT 478
    |||:::|||||:::|||:::|||:::|||:::|||:::|||:::|||:::|||
QY 328 IleMetAlaLysValTyrValLeuAlaSerGlyAlaIleGlyAsnProGlnIleLeuTyr 347
    |||:::|||||:::|||:::|||:::|||:::|||:::|||:::|||:::|||
Db 479 CTACTTCT-----AATGATTTCATGCTCTGCT 505
    |||:::|||||:::|||:::|||:::|||:::|||:::|||:::|||:::|||
QY 348 AsnSerGlyPheSerGlyLeuGlnValThrProArgAsnAspSerLeuIleProAsnLeu 367
    |||:::|||||:::|||:::|||:::|||:::|||:::|||:::|||:::|||
Db 506 CATCTCTGTGTTTACGGGC-----523
    |||:::|||||:::|||:::|||:::|||:::|||:::|||:::|||:::|||
QY 368 GlyArgTyrIleThrGluGlnProMetAlaPheCysGlnIleValLeuArgGlnGluPhe 387
    |||:::|||||:::|||:::|||:::|||:::|||:::|||:::|||:::|||
Db 523 -----523
    |||:::|||||:::|||:::|||:::|||:::|||:::|||:::|||:::|||
QY 388 ValAspSerValArgAspProTyrGlyLeuProTrpTrpLysGluAlaValAlaGln 407
    |||:::|||||:::|||:::|||:::|||:::|||:::|||:::|||:::|||
Db 523 -----523
    |||:::|||||:::|||:::|||:::|||:::|||:::|||:::|||:::|||
QY 408 HisIleAlaLysAsnProThrAspAlaLeuProIleProPheArgAspProGluProGln 427
    |||:::|||||:::|||:::|||:::|||:::|||:::|||:::|||:::|||
Db 524 CATCTATAGACTTAGTGTCTACTCTTGCACCACCTCCCTTA-----CAACTTGAA 574
    |||:::|||||:::|||:::|||:::|||:::|||:::|||:::|||:::|||
QY 428 ValThrThrProPheThrGluGluHis-----ProTrpHisThrGlnIleHisArgAsp 445
    |||:::|||||:::|||:::|||:::|||:::|||:::|||:::|||:::|||
Db 575 ATTCTCTCACCCTTAACATCAGAACATCAATTTCTAGTGGGAACAGCGTTCCCATTTACAG 634
    |||:::|||||:::|||:::|||:::|||:::|||:::|||:::|||:::|||
QY 446 AlaPheSer-----TyrGlyAlaValGlyProGluValAspSer 458
    |||:::|||||:::|||:::|||:::|||:::|||:::|||:::|||:::|||
Db 635 GCCTTCTCTCCAACATTTTCACCGGAGCGCTGCACCCGAGGTGTATACC 682
    |||:::|||||:::|||:::|||:::|||:::|||:::|||:::|||:::|||

RESULT 13
BH494670/c
LOCUS BH494670
DEFINITION BH494670 BOGJ Brassica oleracea genomic clone BOGJV59, genomic
survey sequence.
ACCESSION BH494670
VERSION BH494670.1 GI:17702774
```


Alignment Scores:

Pred. No.:	9.26	Length:	920
Score:	106.00	Matches:	79
Percent Similarity:	34.78%	Conservative:	41
Best Local Similarity:	22.90%	Mismatches:	128
Query Match:	3.23%	Indels:	98
DB:	10	Gaps:	12

US-09-856-327-2 (1-618) x BE782505 (1-920)

QY	123	AsnValProThrLeuAspProGlyAlaTrpSerAlaProProGlySerSerAlaIleSer	142
DB	846	CATATCCAGACTATCAAGAGGTCTAAGCCAGCTCCACCTTTGTCTCTTGATAAA	787
QY	143	AsnGlyLysAsnProHisGlnArgGluPheGluAsnLeuSer	156
DB	786	AATGTAGACTTCCAAATTTGAGCCGAGCTGTATTCATCTTTGGGCCACATTTCTCT	727
QY	157	---AlaGluAlaValThrArgGlyValGlyGly-MetSerThrHisTrpThrCysSerTh	175
DB	726	GGGTCCAAAGGCTCTCGCACCTACTTCTGTGCTCTCAGGTTTCTGCCACTTGTCCAC	667
QY	175	rProArgIleHisProPrometGluSerLeuProGlyIleGlyArgProLysLeuSerAs	195
DB	666	TCCTCTTGTGTTT-----ATTCAACCTCCGCAATATATCTCA	631
QY	195	nAspProAla-GluAspAspLysGluTrpAsnGluLeuTySerGluAlaGluArgLeuI	215
DB	630	GGAAAGTGCTGAA-----TGTCGTCTCATTCACGAGTAG	595
QY	215	leGlyThrSerThr-----LysGluPhe-AspGluSerIleArgHisThrLeuValLeu	232
DB	594	TTTCAACAAGAGTGTCTTCTCATTTTCAGATTCTCGGTACCACATCTTTGGGGCTA	535
QY	233	ArgSerLeuGlnAspAlaTyrlsAspArgGlnArgIlePheArgProLeuAla	252
DB	534	AG-----GTATCACCCGAAAGCCCAATGATGCTCCGCAATTAATGTTGC	487
QY	253	CysHisArgLeuLysAsn-AlaProGluTrpValGluTrpHisSerAlaGluAsnLeuPh	272
DB	486	TGTCACGCGCGGTAAATCATACAGAAATATGTTCTCTGTGCACAGGCTTCGGCCCAAC	427
QY	272	eHisSerIleTyrlsAsnAspLysGlnLysLysLeuPheThrLeuLeuThrAsnHisAr	292
DB	426	AAATCCGCTTAC-----AGGCAGGAGCTTCACAGAAATCGG-ACAAACACAC--	382
QY	292	gCysThrArgLeuAlaLeuThrGlyGlyTyrlsGlyLysLysIleGlyAlaAlaGluValAr	312
DB	381	TGTACTCGAGCAGCTTCTTAGGAGGAGGATCAGAG-----	346
QY	312	gAsnLeuAlaThrArgAsnProSerSerGlnLeuAspSerTyrlleMetAlaLysVa	332
DB	345	-----TCATCTTCGACAACTCAAGCAATTTAGCACATAC-----	310
QY	332	lTyrlValLeuAlaSerGlyAlaIleGlyAsnProGlnIleLeuTyrlAsnSerGlyPheSe	352
DB	309	-----GGGTTTTC	302
QY	352	rglyLeuGlnValThrProArgAsnAspSerLeuIleProAsnLeuGlyArgTyrlleTh	372
DB	301	A-----TCATCATCCCTTCAATCAACACAACTGTCCAAATCTGGATGTGAATCTCGGT	248
QY	372	rgluGlnProMetAlaPheCysGlnIleValLeuArgGln-----GluPheValAs	389
DB	247	GGAAACACCTTCTGTTTTCACACACATTTCTCTATAGTTTGGTAGTGCAGTTTTCGATC	188
QY	389	pSerValArgAspProTyrlGlyLeuProTrpTrpLysGluAlaVal-----	405
DB	187	CACAAGGCGCTGCCAACCCATGAATAAGTTTCTTCTGCTTTCAGCCTTGTGGGTAGTGT	128
QY	405	-----AlaGlnHi	408
DB	127	GCCATGCCGGCTCCAAAGCTTCTCCGGCTAGCCGCCAAGCCCTCCGTTTCTTACCGCCTC	68

QY 408 sileAlaLysAsnProThrAspAlaLeuProfileProPheArgAspProGluProGlnVa 428
DB 67 CACTTCGAGAGTCGCGGGGATCGCGACCTCTAACCTGCACAGGCTCTCTCTAC 8
QY 428 lThrThr 430
DB 7 AACTACG 1

RESULT 15
BC034946 4866 bp mRNA linear HTC 04-MAR-2003
LOCUS Homo sapiens, Similar to secretory pathway component Sec31B-1,
DEFINITION clone IMAGE:4826447, mRNA.
ACCESSION BC034946
VERSION BC034946.1 GI:23273821
KEYWORDS HTC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 4866)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (31-JUL-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Miklos Paikovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Institute for Systems Biology
contact: amadan@systemsbiology.org
Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha
Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 33 Row: 0 Column: 2
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 14591929
This clone has the following problem: retained intron.

FEATURES
source

1..4866
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4826447"
/tissue_type="Testis"
/clone_lib="NIH_MGC_97"
/lab_host="DH10B"
/note="Vector: pBluescript"

BASE COUNT 1104 a 1289 c 1294 g 1179 t
ORIGIN

Alignment Scores:

Pred. No.:	91	Length:	4866
Score:	106.00	Matches:	81
Percent Similarity:	34.05%	Conservative:	46
Best Local Similarity:	21.72%	Mismatches:	137
Query Match:	3.23%	Indels:	109
DB:	11	Gaps:	18

US-09-856-327-2 (1-618) x BC034946 (1-4866)

QY	229	ThrLeuValLeuArgSerLeuGlnAspAlaTyrlsAspArgGlnArgIlePheArgPro	248
DB	2459	ACAGATCTGCTGAAGCAACACACAGGCGCTACTTGGCCCAAGAGAAACCAAAATCTCC	2518

Search completed: September 24, 2003, 08:50:25
Job time : 3544 secs

```
Qy 249 LeuProLeuAlaCysHisArgLeuLysAsnAlaProGluTyrVal----- 263
Db      |||||
Db 2519 TCGCTTCTAGCTGTGTGTCACAAAGAAATGGGAAGGATGGTGTGTACCTGTAGCCTG 2578
Qy 264 ---GluTrpHisSerAlaGluAsnLeuPheHisSerIleTyrAsnAspAspLysGlnLys 282
Db      |||||
Db 2579 AAGAACTGGAGAGGAGCTGGCTTGTCTACTGACATACTACGACACAGAGAAATTTCCC 2638
Qy 283 LysLeuPheThrLeuThrAsnHisArgCysThrArgLeu----- 296
Db      |||||
Db 2639 GAGCTCTGTGCATGCTGGGA-----ACTGCGATGGGAACAGGAGGGGAGCAGG 2686
Qy 297 AlaLeuThr-----GlyGlyTyrGluLysLysIle 306
Db      |||||
Db 2687 GCACCTAACCTCCGAAGCCAGAGCTCTGTATGTGTGCTCAGGAGGTGGAGCGCTGGTG 2746
Qy 307 GlyAla---AlaGluValArgAsnLeuLeuAlaThrArgAsnProSerSerGlnLeuAsp 325
Db      |||||
Db 2747 GAGTGTGGGGCAAAATGCCACGAGCTTTG-----TCCCCATGGCTCTGCAGGAC 2797
Qy 326 SerTyrIleMetAlaLysValTyrValLeuAlaSerGlyAlaIleGlyAsnProGlnIle 345
Db      |||||
Db 2798 -----CTGATGGAGAGGTGATGTTCTT-----AACAGGAGCTTG 2833
Qy 346 LeuTyrAsnSerGlyPheSerGlyLeuGlnValThrProArgAsnAspSerLeuIlePro 365
Db      |||||
Db 2834 GAGCAACTGCGGGGTCTCATGTGGGTGAGCCAGCCCTGCCACAACCTACAGGGTCACT 2893
Qy 366 AsnLeuGlyArgTyrIleThrGluGln-----PrometAlaPheCysGln 380
Db      |||||
Db 2894 CAGTATGCCAACCTCTCTGGCAGCCAGCCAGCCAGCTGGCCACTGCCATGAGCTTCTACCC 2953
Qy 381 IleValLeuArgGlnGluPheValAspSerValArgAsp----- 393
Db      |||||
Db 2954 AGGCACTGTGCTCAGCCACCAGTTACGAGCTAAGAGATCGGCTTTTCATGCTCAAGGT 3013
Qy 394 -----AspProTyrGlyLeuPro----- 399
Db      |||||
Db 3014 TCTGTGTCTTGGGCCAACAGTCTCCCGCTTCCCTTCCCGCGATTGTTGGGAGCT 3073
Qy 400 -----TrpTrpLysGluAlaValAlaGlnHisIleAlaLysAsnProThrAspAlaLeu 417
Db      |||||
Db 3074 ACCCTCCACTTAAGAGACATCATCTACAGATTGGGATCCCGCCCTTCTCACCAGGTC 3133
Qy 418 ProIleProPheArgAspProGluProGlnValThrProPheThr-GluGluHis-- 436
Db      |||||
Db 3134 CCAACTCCA-----TCTCCAAGCCCAAGGTTTTTACCCCTCAGTCACTCACCAGCGATG 3187
Qy 437 -ProTrpHisThrGlnIle-----HisArgAspAlaPheSe 448
Db      |||||
Db 3188 CCCTTGGCACCTTCCCATCTCTAGCCCTTATCAGGGTCCCGAGGACACAGATATAAGTGAC 3247
Qy 448 rTyrGlyAlaValGlyProGluValAspSerArgValIleValAspLeuArgTrpPheG 468
Db      |||||
Db 3248 TACAGGGCACCTGGGGCCCCAG----- 3269
Qy 468 yAlaThrAspProGluAlaAsnAsnLeuLeuValPheGlnAsnAspValGlnAspGlyTy 488
Db      |||||
Db 3270 -GCCATCCAGCT-----TTGCTTTTGGCCCTTGGGTTAAGGCTGGA-- 3311
Qy 488 rSerMetProGlnProThrPheArgTyrArgProSerThrAlaSerAsnValArgAlaAr 508
Db      |||||
Db 3312 -----CCTCAAGATTCCTGGAAGAAGCCCGCCAGCCCGGGAACCTCCAGAGGAA 3364
Qy 508 gLysMet-----MetAlaAspMetCysGluValAl 518
Db      |||||
Db 3365 CAAGCTGCAGAGACATTTATGCCCGCAGCACCAATTACTCTCCAGTTATGAGCCTCAC 3424
Qy 518 aSerAsnLeuGlyTyrLeuProThrSerProPro 530
Db      |||||
Db 3425 CCCTGAGCTACAGGGATTTCTTCCTCAGACCCCCCT 3461
```

This Page Blank (uspto)

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	100	100	100	1	...

1	3284	100.0	2106	22	AA999980	Nucleotide sequ
2	1232	37.5	1902	21	AA246411	Pleurotus cornuc
3	1222.5	37.2	1869	17	AA344420	Pyranose oxidase
4	1210	36.8	1869	22	AA887519	Trametes hirsuta
5	1210	36.8	1895	22	AA887518	Trametes hirsuta
6	1091	33.2	1701	20	AAV83626	Nucleic acid en
7	1044	31.8	1946	21	AAAF7187	T matsutake pvt

Accession number	Gene	Accession number	Gene
8	1044	9	208.5
8	31.8	9	6.3
8	1946	9	2467
8	21	9	24
8	AAAU/403	9	ABK52367

9	206.3	0.5	2467	24	ABX52307	DNA encoding gly-
10	207	6.3	4290	23	ABL52917	2-keto-D-glucos-
11	206	6.3	5187	20	AAK57912	G. oxydans D-son-
12	200	6.1	1632	20	AAK57909	G. oxydans D-son-
13	198	6.0	4665	20	AAK25345	Membrane-bound

Accession	Year	Location	Host	Pathogen		
14	1989	24	1689	24	ABL52/44	Glucose-3-denitrification
15	1995	21	2484	21	AAC36995	Arabidopsis thaliana

c	16	148.5	4.5	1641	20	AAZ10823	Choline oxidase
	17	148	4.5	4403765	22	AAI99683	Mycobacterium tuberculosis

C	19	124	3.8	2400	18	AA75000	Choline oxidase
	18	148	4.3	4411529	22	AA199682	Mycobacterium
	20	124	3.8	2576	21	AA295701	Humicola insolens

21	124	2576	22	AAH47743	H. insolens DSM
22	122.5	1503	22	AAF32484	A. tumefaciens 1

22	122.5	3.7	1303	22	AAT32404	A. tumefaciens 1
23	122.5	3.7	1841	22	AAF32462	A. tumefaciens 1
24	121.5	3.7	2400	17	AAT42859	Choline oxidase

26	121	3.7	2944528	24	ABA03041	<i>Listeria monocytogenes</i>
25	121	3.7	4233	20	AAX87946	<i>Candida tropicalis</i>

27	120.5	3.7	8918	21	AA363625	Nucleotide sequence
28	120.5	3.7	9199	20	AA384322	Stealth virus nucleotide sequence
29	119	3.6	2336	21	AAC44859	Arabidopsis thaliana
30	119	3.6	8526	23	AA579203	DNA encoding novel protein
31	116.5	3.5	4305	20	AA387945	Candida cloacae
32	115.5	3.5	29879	14	AA046806	eryA region of Salmonella
33	115	3.5	1596	22	AA60988	P. putida KT2440
34	114	3.5	4403765	22	AA199683	Mycobacterium tuberculosis
35	114	3.5	441529	22	AA199682	Mycobacterium tuberculosis
36	113.5	3.5	35829	23	AA559573	Propionibacterium freudenreichii

37	113	3.4	6233	24	ABQ710066	<i>Listeria monocytogenes</i>	
C	38	113	3.4	21185	21	AA663350	<i>Streptomyces glaucus</i>
C	39	113	3.4	63164	21	AA663348	<i>Streptomyces glaucus</i>
C	40	112.5	3.4	2385	13	AA027956	<i>Cholesterol oxidase</i>
C	41	112	3.4	1911	25	ABX56251	CDNA encoding coagulase
C	42	111	3.4	29379	23	AA559510	<i>Propionibacterium freudenreichii</i>
C	43	111	3.4	53179	23	AA559543	<i>Propionibacterium freudenreichii</i>
C	44	110.5	3.4	34980	24	AB081848	<i>Bifidobacterium bifidum</i>
C	45	109.5	3.3	1818	21	AA248464	Glucose oxidase

		ALIGNMENTS	
RESULT	1		
AAF99980			
ID	AAF99980	standard;	CDNA; 2106 BP.
XX			
XX	AAF99980;		
XX			
DT	20-JUL-2001	(first entry)	
XX			
DE	Nucleotide sequence encoding <i>Lyophyllum shimeji</i> antibacterial protein.		
XX			
XX	Fungus; antibacterial; antibiotic; plant pathogen; bacterial infection;		
KW	<i>Pycularia orizae</i> ; <i>Rhizoctonia solani</i> ; rice pathogen; ss.		
XX			
OS	<i>Lyophyllum shimeji</i> .		

AA	
FH	Key
FT	Location/Qualifiers
	8..1864
	CDS

FT /*tag= a
 FT /product= "antibacterial protein"

PN W0200121657-A1.

PD 29-MAR-2001.

PF 20-SEP-2000; 2000WO-JP06404.

PR 21-SEP-1999; 95JP-0267238.

PA (NISB) JAPAN TOBACCO INC.

PA (NORQ) SOC TECHNO-INNOVATION AGRIC FORESTY & FI.

PI Takakura Y, Kuwata S, Inoue Y;

DR WPI; 2001-281598/29.

DR P-PSDB; AAB97035.

PT Antibacterial protein and encoded gene isolated from *Lyophyllum*
 PT *shimeji*, with activity against plant pathogenic bacteria, applicable in
 PT agriculture e.g. rice cultivation at low concentration, produced at low
 PT cost on large scale .

PS Claim 12; Page 38-42; 52pp; Japanese.

CC The present sequence encodes an antibacterial protein from the fungus
 CC *Lyophyllum shimeji*. The protein was obtained from a fraction prepared
 CC by extracting *Lyophyllum shimeji* with water and subjecting the extract
 CC to ammonium sulphate precipitation. The protein inhibits the growth
 CC of the plant pathogenic bacteria *Pycularia oryzae* and *Rhizoctonia*
 CC *solani* at a relatively low concentration. *P. oryzae* and *R. solani* are
 CC causative of the two major diseases of rice. The protein contains
 CC components of 70 kDa and 65 kDa, as determined by SDS-PAGE.

CC The antibacterial protein can be produced at low cost on a large scale.

SQ Sequence 2106 BP; 593 A; 540 C; 527 G; 446 T; 0 other;

Alignment Scores:

Pred. No.: 2,066-279 Length: 2106
 Score: 3284.00 Matches: 618
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 22 Gaps: 0

US-09-856-327-2 (1-618) x AAF99980 (1-2106)

QY 1 MetSerLeuSerThrGluGlnMetLeuArgAspTyrProArgSerMetGlnIleAsnGly 20
 DB 8 ATGTCTCTCAACCGAGCAGATGCTACGCGACTATCCACGGTCTATGCAAAATCAACGGA 67
 QY 21 GlnIleProLysAsnAlaIleHisGluThrTyrGlyAsnAspGlyValAspValPheIle 40
 DB 68 CAGATTCTTAAGACGCAATTCACGAAACATACGGAACACGACGAGGTGATGTATTCATT 127
 QY 41 AlaGlySerGlyProIleGlyAlaThrTyrAlaLysLeuCysValGluAlaGlyLeuArg 60
 DB 128 GCAGGATCTGCACCCATTTGGAGCAGCATATCGAAAGCTCTGTGTGAAGCTGGTCTACGT 187
 QY 61 ValValMetValGluIleGlyAlaAlaAspSerPheTyrAlaValAsnAlaGluGly 80
 DB 188 GTTGTGATGGTCGAGATCGGAGCTGCTGATAGCTTCTACGCTGTTAATGCGGAAGGGA 247
 QY 81 ThrAlaValProTyrValProGlyTyrHisLysLysAsnGluIleGluPheGlnLysAsp 100
 DB 248 ACTGCAGTTCCTTCAGTCTTCCTGGCTTACCACGAAGAAGATGAATCGAGTTCAGAAAGAT 307
 QY 101 IleAspArgPheValAsnValIleLysGlyAlaLeuGlnGlnValSerValProValArg 120
 DB 308 AFTGACCGCTTCGTCGAATGTAATCAAGGAGCGCTTACAAACAGTCTCTGTTCTCTCAGA 367
 QY 121 AsnGlnAsnValProThrLeuAspProGlyAlaIlePheSerAlaProProGlySerSerAla 140

DB 368 AACCAAGACGTCCTTACACTTGATCCGAGAGCTGGAGCGGCCCTTGAAGTTCAAGCC 427
 QY 141 IleSerAsnGlyLysAsnProHisGlnArgGluPheGluAsnLeuSerAlaGluAlaVal 160
 DB 428 ATATCGAACGGTAAATAATCCCTACCCAGCGGAATTCGAGAACTTGTCTGCGGAGGCCGTA 487
 QY 161 ThrArgGlyValGlyGlyMetSerThrHisTrpThrCysSerThrProArgIleHisPro 180
 DB 488 ACGGTGGAGTCGGGGCATGAGTACCCTCCAGTGGAGTCTCCACGCCAGGATTCATCCA 547
 QY 181 ProMetGluSerLeuProGlyIleGlyArgProLysLeuSerAsnAspProAlaGluAsp 200
 DB 548 CCCATGGAAGTCTCCCGGCATCGCGCTCCGAAGCTCAGTACGACGCCGACAGGAC 607
 QY 201 AspLysGluTrpAsnGluLeuTyrSerGluAlaGluArgLeuIleGlyThrSerThrLys 220
 DB 608 GACAAAGAGTGAACGAGCTTTATTCGAGGCGGAGCTCTCATCGGAGCTTCCACCAAG 667
 QY 221 GluPheAspGluSerIleArgHisThrLeuValLeuArgSerLeuGlnAspAlaTyrLys 240
 DB 668 GAATTCGACGAGTCAATTCGGCACACCCCTTGTCTGCGCTCTTTCGAAGACGGTACAAG 727
 QY 241 AspArgGlnArgIlePheArgProLeuAlaCysHisArgLeuLysAsnAlaPro 260
 DB 728 GATCGTCAACGTATCTTTCGCCCTCTCCCGTTGGCATGCCACCGGTTGAAGAACGCGCG 787
 QY 261 GluTyrValGluTrpHisSerAlaGluAsnLeuPheHisSerIleTyrAsnAspLys 280
 DB 788 GAATACGTCGATGGCATGCACTACAGAAAATCTTTTCCACTCTATCTACAGCATGACAAAG 847
 QY 281 GlnLysLysLeuPheThrLeuLeuThrAsnHisArgCysThrArgLeuAlaLeuThrGly 300
 DB 848 CAGAAGAGCTCTTTACCTCTGCTGACGAACCATCGCTGCACACGACTGCGCTTACGGGC 907
 QY 301 GlyTyrGluLysLysIleGlyAlaAlaGluValArgAsnLeuLeuAlaThrArgAsnPro 320
 DB 908 GGGTATGAGAAAGATTGGCGCTGCCGAGGTCAGGAATCTACTGGCCACCCAGGAATCCT 967
 QY 321 SerSerGlnLeuAspSerTyrIleMetAlaLysValTyrValLeuAlaSerGlyAlaIle 340
 DB 968 AGTTGCGCAGCTGGACAGCTATATATGCGAAGGTATATGACTGGCGCTCGGGAGCGATC 1027
 QY 341 GlyAsnProGlnIleLeuTyrAsnSerGlyPheSerGlyLeuGlnValThrProArgAsn 360
 DB 1028 GGCAACCCACAGATTCTCTATACTCGGCTTCTCTGGGCTACAGGTACGCCACCGCAAT 1087
 QY 361 AspSerLeuIleProAsnLeuGlyArgTyrIleThrGluGlnProMetAlaPheCysGln 380
 DB 1088 GACTCGTTGATCCCAACCTGGGAGGTACATACGAGACGCGACGCGATGGCATTTTGCAG 1147
 QY 381 IleValLeuArgGlnGluPheValAspSerValArgAspAspProTyrGlyLeuProTrp 400
 DB 1148 ATAGCTTGGAGCGAGGAATTCGTGACAGCTGGCGAGCATCTTATGGAGTGGCATGG 1207
 QY 401 TrpLysGluAlaValAlaGlnHisIleAlaLysAsnProThrAspAlaLeuProIlePro 420
 DB 1208 TGGAAAGAGCCGTGTCTCAACATATTCGCAAGAACCCGACAGATGCACCTGCCATTCCG 1267
 QY 421 PheArgAspProGluProGlnValThrThrProPheThrGluGluHisProTrpHisThr 440
 DB 1268 TTCGCGCATCCGGAACCCAGGTAAACCCCATTTACAGGAAGAACACCCCTGCGACAG 1327
 QY 441 GlnIleHisArgAspAlaPheSerTyrGlyAlaValGlyProGluValAspSerArgVal 460
 DB 1328 CAGATTACCCCGATGCTTTTTCGTACGTCGCCGCTCGGTCCTCGAGTGGAGCTCTCGTGT 1387
 QY 461 IleValAspLeuArgTrpPheGlyAlaThrAspProGluAlaAsnAsnLeuValPhe 480
 DB 1388 ATCTGTCGACCTGCGCTGGTTTGGCGCAACCCAGCTGAAGCAACAACTTTTGGTTTC 1447
 QY 481 GlnAsnAspValGlnAspGlyTyrSerMetProGlnProThrPheArgTyrArgProSer 500

Db	1448	CAGNACGATGTTCAAGACGGGTACAGTATGCCGAGCCGACGGTTCAGATATCGACCCAGC	1501
Qy	501	ThrAlaSerAsnValArgAlaArgLysMetMetAlaAspMetCysGluValAlaSerAsn	520
Db	1508	ACTGCGTCAAAACGTGAGAGCAAGGAAAATGATGCCGATATGTGCGAAGTGGCGAGCAAC	1567
Qy	521	LeuGlyGlyTyrLeuProThrSerProGlnPheMetAspProGlyLeuAlaLeuHis	540
Db	1568	TTGGGAGGTATTGTCACAGTCCGCCCACTTTATGGATCATCAGGCGCTTGCACTTCAT	1627
Qy	541	LeuAlaGlyThrThrArgIleGlyPheAspLysAlaThrThrValAlaAspAsnAsnSer	560
Db	1628	CTTGGGGGACTACTCGCATGGCTTCGACAAGCAACTACAGTGGCTGATAACAACCTCG	1687
Qy	561	LeuValTrpAspPheAlaAsnLeuTyrValAlaGlyAsnGlyThrIleArgThrGlyPhe	580
Db	1688	CTGGTCTGGGACTTTGGCAATCTTTATGTTGCAGGCAATGGCACCATCAGGACGGGCTTC	1747
Qy	581	GlyGluAsnProThrLeuThrSerMetCysHisAlaIleLysSerAlaArgSerIleIle	600
Db	1748	GCGGAGAACCACCGACACTACGTGCGATGTGCCAGCTATCAAGAGCGGAGGAGCATCATC	1807
Qy	601	AsnThrLeuLysGlyGlyThrAspGlyLysAsnThrGlyGluHisArgAsnLeu	618
Db	1808	AATACACTCAAGGTGGGACTGACGGGAAAAATACAGGCGAGCATCGCAACCTT	1861
RESULT 2			
AAZ46411			
ID AAZ46411 standard; DNA; 1902 BP.			
XX	AAZ46411;		
AC	AAZ46411;		
XX			
XX	07-MAR-2000 (first entry)		
XX			
DE	Pleurotus cornucopiae antitumour protein coding sequence.		
XX			
KW	Antitumour; cancer; tumour; treatment; expression; tumour suppressor;		
KW	p53; pBR; ss.		
XX			
OS	Pleurotus cornucopiae.		
XX			
PN	JPl1315096-A.		
XX			
PD	16-NOV-1999.		
XX			
PF	07-AUG-1998; 98JP-0236349.		
XX			
PR	08-AUG-1997; 97JP-0215311.		
PR	02-MAR-1998; 98JP-0066176.		
XX			
PA	(NEWF-) NEW FOOD CREATION GIJUTSU KENKYU KUMIAI.		
XX			
XX	WPI: 2000-058170/05.		
DR	P-PSDB; AAY52700.		
DR			
XX			
PT	An antitumour protein derived from Pleurotus cornucopiae and its gene -		
PT	useful for treatment of cancer including those caused by abnormal		
PT	expression of cancer inhibitory gene (e.g. p53 and pBR)		
XX			
PS	Claim 9; Page 15-16; 23pp; Japanese.		
XX			
CC	The invention relates to a novel antitumour protein extracted from		
CC	fruiting bodies of the fungus Pleurotus cornucopiae. The protein and		
CC	nucleotides encoding it are useful for the treatment of cancer,		
CC	including those caused by abnormal expression of tumour suppressor		
CC	genes such as p53 and pBR. This sequence represents the coding		
CC	sequence of the gene encoding the antitumour protein.		
XX			
SQ	Sequence 1902 BP; 437 A; 556 C; 479 G; 430 T; 0 other;		
Alignment Scores:			
Pred. No.:	1.18e-98	Length:	1902
Score:	1232.00	Matches:	275

Percent Similarity:	57.05%	Conservative:	77		
Best Local Similarity:	44.57%	Mismatches:	193		
Query Match:	37.52%	Indels:	72		
DB:	21	Gaps:	18		
<hr/>					
US-09-056-327-2 (1-618) x AAZ46411 (1-1902)					
Qy	18	IleAsnGlyGlnIleProLysAsnAlaIleHisGluThrTyrGlyAsnAspGlyValAsp	37		
Db	130	ATTCGGGAACACTCCGTGGGA-----GATACATCCAAGTTTCGAC	171		
Qy	38	valPheIlealaGlySerGlyProIleGlyAlaThrTyrAlaLysLeuCysValGluAla	57		
Db	172	GTTGTCACTCGTCGGCTGTGCCAGATTGGCTTACTATGCCTCCTCGTCGAGGCA	231		
Qy	58	GlyLeuArgValValMetValGluIleGlyAlaAlaAspSerPheTyrAlaValAsnAla	77		
Db	232	GGTTTCAAATAGTAGTGTGAAATGGAGAATCGACTCT-----	273		
Qy	78	GluGluGlyThrAlaValProTyrValProGlyTyrHisLysAsnGluIleGluPhe	97		
Db	274	-----GGCGCGAAGCTT-----GGITCCCATAAGAAAAATACGGTTGAGTAC	315		
Qy	98	GlnLysAspIleaspArgPheValAsnValIleLysGlyAlaLeuGlnValSerVal	117		
Db	316	CAGAAGAACAATTGCACAAATTCGTCATGTCTATCAAGGACACACTCATCGCTGATCCGTC	375		
Qy	118	ProValArgAsnGlnAsnValProThrLeuAspProGlyAlaTrpSerAlaProGly	137		
Db	376	CCGTGTCAACAAATATGTTGCCGATACTCTCAGCCCCGCTAGTCGGAGCGCC-----TCA	429		
Qy	138	SerSerAlaIleSerAsnGlyLysAsnProHisGlnArgGluPheGluAsnLeuSerAla	157		
Db	430	ACACACTTTGTCGCGAACGGCGCAACCCGACCAAGATCCCTTTACTATCTCTCAGGC	489		
Qy	158	GluAlaValThrArgGlyValGlyGlyMetSerThrHisThrPyrCysSerThrProArg	177		
Db	490	CAAGCTGTGACC CGCTGTGTGGAGGAATGGCGACGCACTGGACTTGCGCAACACCTCGC	549		
Qy	178	IleHisProProMetGluSerLeuProGlyIleGlyArgProLysLeu---SerAsnAsp	196		
Db	550	TTCACAAGAGT-----GAGCGGCTTAAGCTTTGTGAAGATGAT	588		
Qy	197	ProAlaGluAspAspLysGluThrAsnGluLeuTyrSerGluAlaGluArgLeuIleGly	216		
Db	589	GACTCCGGGACCCCATCGAGTGGGAACGCTCTATGCATATCCCGAATCGTTCGTC AAG	648		
Qy	217	ThrSerThrLysGluPheAspGluSerIleArgHisThrLeuValLeuArgSerLeuGln	236		
Db	649	ACCGGGCATAAATCAGTTTGATCAATCTATCCGCAACACCTTGCTACTCGAAGAGCTTCAG	708		
Qy	237	AspAlaTyrLysAspArgGlnArgIlePheArgProLeuAlacysHisArgLeu	256		
Db	709	GAGTCTCATCAGGT---CAGCCGGATTGTAGCAGATCCCTCTCGGCCCAACGCAACC	765		
Qy	257	LysAsnAlaProGluTyrValGluThrHisSerAlaGluAsnLeuPhe-----HisSer	274		
Db	766	AAC-----CCCGCTTCGTGAGTGGAGCTTCGCGCATACGGTCTTTGNACCTTGAGAAC	819		
Qy	275	IleTyrAsnAspAspLysGlnLysLysLeuPheThrLeuThrAsnHisArgCysThr	294		
Db	820	CGCCCCAACCGCATGACGAGAGAGGGCGTTTTCAACCTCTTCCCGCACGTTGATGTGAA	879		
Qy	295	ArgLeuAlaLeuThrGlyGlyTyrGluLysLysIleGlyAlaAlaGluValArgAsnLeu	314		
Db	880	CGCGTTACG---CGGGATTCCCTCGACCGGAAGATAGAGGACATTTGAGTCCATGACCTG	936		
Qy	315	LeuAlaThrArgAsnProSerSerGlnLeuAspSerTyr--ileMetAlaLysValTyr	333		
Db	937	ATTAGCGCGC-----GATFCGCTACAAGGTCAAGGCTGACGTTGTTT	975		
Qy	334	ValLeuAlaSerGlyAlaIleGlyAsnProGlnIleLeuTyrAsnSerGlyPheSerGly	353		


```
Db 520 CGTTTGTACCGGAGCAG-----CGCCCGTTGCTCGTGAAGAC 558
Qy 197 ProAlaGlu----AspAspLysGluTrpAsnGluLeuTyrSerGluAlaGluArgLeuIle 215
Db 559 GACCAGGACGCTGACGAGCGCGAGTGGAGCGGCTGTACACCAAGCGCGATCATACTTC 618
Qy 216 GlyThrSerThrLysGluPheAspGluSerIleArgHisThrLeuValLeuLeuArgSerLeu 235
Db 619 AAGACCGCGGACCGACCGTTCAGGAGTTCGATCCGCCACACACCTCGTCTCAACAAGTTC 678
Qy 236 GlnAspAlaTyrLysAspArgGlnArgIlePheArgProLeuProLeuAlaCysHisArg 255
Db 679 GCGGAGGAACAAGGT---CAGCGGACACTCCAGCAGATCCCCCTCGCGCAACGCGT 735
Qy 256 LeuLysAsnAlaProGluTyrValIleTrpHisSerAlaGluAsnLeuPhe-----His 273
Db 736 -----CGCAGCTCCGACCTTCGTCGAGTGGAGCTCGGGAACACCGTGTTCGACCTCCAG 789
Qy 274 SerIleTyrAsnAspAspLysGlnLysLysLeuPheThrLeuLeuThrAsnHisArgCys 293
Db 790 AACAGCGCGAACAGCGGACCGCGCAATGAGCGCTTCAACCTCTTCCCGCGGTTGCATGT 849
Qy 294 ThrArgLeuAlaLeuThrGlyGlyTyrGluLysLysIleGlyAlaAlaGluValArgAsn 313
Db 850 GAGCGCGTC-----GTGCGCAAC 867
Qy 314 LeuLeuAlaThrArgAsnProSerSerGlnLeuAspSertyr----- 327
Db 868 -----ACGTCAACTCCGAGATCGAGAGTCTGCACATCCACGACCTCATC 912
Qy 328 -----IleMetAlaLysValTyrValLeuAlaSerGlyAlaIleGly 341
Db 913 TCCGCGCACCGCTTCGAAATCAAAAGCAGACGTGTTCGTCTTACAGCGCGGCGGTCCAC 972
Qy 342 AsnProGlnIleLeuTyrAsnSerGlyPheSerGlyLeu---GlnValThrProArgAsn 360
Db 973 AACCGCGAGCTTCGTGAACTTCGTGGCTTTCGACAGCTGGCGCGCGGACCCCGCGAAC 1032
Qy 361 ---AspSerLeuIleProAsnLeuGlyArgTyrIleThrGluGlnProMetAlaPheCys 379
Db 1033 CCGCGCGAGTTGCTGCGCTCCCTCGGAAGTATACATCCAGAGAGTCTGCTCTTCTGC 1092
Qy 380 GlnIleValLeuArgGlnGluPheValAspSerValArgAspAspProTyr-----Gly 397
Db 1093 CAGACCGTGTATGAGCAGCGAGCTCATCAGACGCTCAAGTCCGACATCATCAGGGGC 1152
Qy 398 LeuPro----- 399
Db 1153 AACCCCTGGCGATCTGGGGTACAGCGTACAGCTACAGCCCGCGCGGAGACCAACAAGCAC 1212
Qy 400 -----TrpTrpLysGluAlaValAlaGlnHisIleAlaLysAsnProThrAspAlaLeu 417
Db 1213 CCGGACTGGTGGAAAGGTGAAGACCATGATGACAGCAGCAGGAGGACCCGCTT 1272
Qy 418 ProfileProPheArgAspProGluProGlnValThrThrProPheThrGluGluHisPro 437
Db 1273 CCAATCCGTTTCGAGGACCCGAGCCGAGTCCACCTTGTTCACGCCATCCACCCG 1332
Qy 438 TrpHisThrGlnIleHisArgAspAlaPheSerTyrGlyAlaValGlyProGluValAsp 457
Db 1333 TGGCACACTCAGATACCGCGATGCGTTTACGTCGCGCGGTTCAGCAACACCTGAC 1392
Qy 458 SerArgValIleValAspLeuArgTrpPheGlyAlaThrAspProGluAlaAsnAspLeu 477
Db 1393 TCAGCTCTCATCGTGCAGTGGCGTTCCTTGGCGGAGGAGGCCAAGAGGAGGACAACAG 1452
Qy 478 LeuValPheGlnAsnAspValGlnAspGlyTyrSerMetProGlnProThrPheArgTyr 497
Db 1453 CTCCTGGTTCGCGACAAAATTACGACACGCTTACAACATCGCGACGCGAGCTTCGACTTC 1512
Qy 498 Arg---ProSerThrAlaSerAsnValArgAlaArgLysMetMetAlaAspMetCysGlu 516
Db 1513 CGTTCCTCCGGCGCGCCAGCAGCAGGAGCGGAGGACATGATGACCGATATGTGCGTT 1572
```

```
Qy 517 ValAlaSerAsnLeuGlyGlyTyrLeuProThrSerProGlnPheMetAspProGly 536
Db 1573 ATGTCCGGGAAGATTGGTGGCTCTCTGCGCGGCTCCCTCCCGCAATTATGAGGCCGCT 1632
Qy 537 LeuAlaLeuHisLeuAlaGlyThrThrArgIleGlyPheAspLysAla-----ThrThr 554
Db 1633 CTTGTCTCTTCACTCGGTGGTACGCACGCGATGGCTTCGACGAGCAGGAGCAAGTGC 1692
Qy 555 ValAlaAspAsnAsnSerLeuValTrpAspPheAlaAsnLeuTyrValAlaGlyAsnGly 574
Db 1693 TCGGTCAACACGGACTCCGCGGTGTTCGCTTCAAGAACCTGTCTCGTGGCTGCGGA 1752
Qy 575 ThrIleArgThrGlyPheGlyGluAsnProThrLeuThrSerMetCysHisAlaIleLys 594
Db 1753 AACATTCACCCGCGTACGCGCGCAACCCGAGCTCACCGCAATGCTCGCTCGGATCAAG 1812
Qy 595 SerAlaArgSerIleIleAsn 601
Db 1813 AGTTGCGAGTACATCAAGAAC 1833
RESULT 4
AAC87519
ID AAC87519 standard; DNA; 1869 BP.
AC AAC87519;
XX
DT 13-MAR-2001 (first entry)
XX
DE Trametes hirsuta pyranose oxidase cDNA, SEQ ID NO:1 (version 2).
XX
KW Pyranose oxidase; expression construct; recombinant production;
KW monosaccharide oxidation; 2-keto derivative;
KW hydrogen peroxide production; ss.
XX
OS Trametes hirsuta.
XX
PN US6146865-A.
XX
PD 14-NOV-2000.
XX
PF 05-MAY-1999; 99US-0305381.
XX
PR 08-JUN-1998; 98DK-0000774.
XX
PR 10-JUN-1998; 98US-0088724.
XX
PA (NOVO ) NOVO NORDISK AS.
XX
PI Schneider P, Christensen S, Lassen SF;
XX
WP1; 2001-049055/06.
XX
P-PSDB; AAB48832.
XX
PT Novel nucleic acid molecule encoding polypeptide having pyranose
PT oxidase activity used to design oligonucleotide probes to identify and
PT clone DNA encoding the polypeptide from different genera or species -
XX
PS Claim 2; Column 25-28; 20pp; English.
XX
XX The invention relates to nucleic acids (e.g., AAC87518, AAC87519) which
XX encode Trametes hirsuta pyranose oxidase (AAB48832). The invention also
XX relates to expression constructs, expression vectors and recombinant
XX cells comprising pyranose oxidase nucleic acid sequences, and the
XX recombinant production of Trametes hirsuta pyranose oxidase. Pyranose
XX oxidase catalyses the oxidation of several monosaccharides in the
XX pyranose form at position C2 to produce 2-keto derivatives with the
XX release of hydrogen peroxide. Nucleic acids encoding Trametes hirsuta
XX pyranose oxidase may be used to produce the enzyme and to design
XX oligonucleotide probes to identify and clone genomic pyranose oxidase
XX cDNA or genomic DNA from different genera or species of microorganisms
XX (fungi or bacteria). The present sequence represents a cDNA encoding
XX pyranose oxidase from the fungus Trametes hirsuta.
XX Note: Both AAC87518 and AAC87519 are Trametes hirsuta pyranose oxidase
```

CC cDNA sequences which contain the entire open reading frame (ORF).
 CC However, the two sequences have different stop codons - AAC87518 has an
 CC opal stop codon while AAC87519 has an amber stop codon.
 XX
 SQ Sequence 1869 BP; 397 A; 622 C; 548 G; 302 T; 0 other;

Alignment Scores:

Pred. No.:	9,98e-97	Length:	1869
Score:	1210.00	Matches:	274
Percent Similarity:	56.19%	Conservative:	71
Best Local Similarity:	44.63%	Mismatches:	197
Query Match:	36.85%	Indels:	72
DB:	22	Gaps:	16

US-09-856-327-2 (1-618) x AAC87519 (1-1869)

QY	37	AspValPheIleAlaGlySerGlyProIleGlyAlaThrTyrAlaLysLeuValGlu	56
DB	142	GACGTGCCATCTCGGCTCTGGCGGATTGGTCACATATGCGCGGAGTCGTGTGAG	201
QY	57	AlaGlyLeuArgValValMetValGluIleGlyAlaAlaAspSerPheTyrAlaValAsn	76
DB	202	GCGGCTTCAACGTCGCCATGTCGAGATTGGAGATCGACTCCGGCTTGAAGATC---	258
QY	77	AlaGluGluGlyThrAlaValProTyrValProGlyTyrHisLysLysAsnGluIleGlu	96
DB	259	-----GGCTCACACAAGAAGAACACCGTCGAG	285
QY	97	PheGlnLysAspIleAspArgPheValAsnValIleLysGlyAlaLeuGlnValSer	116
DB	286	TACCAGAAGACATCGACAATTCGTAAATGTTATACAAGGGCAACTTATGCCCGTCTCG	345
QY	117	ValProValArgAsnGlnAsnValProThrLeuAspProGlyAlaThrPheSerAlaPro	136
DB	346	GTGCCGTCACACGATGTCGTGTGACACGCTAAGCCCGCGCTCATGGCAAGCT-----	399
QY	137	GlySerSerAlaIleSerAsnGlyLysAsnProHisGlnArgGluPheGluAsnLeuSer	156
DB	400	TCGACGTTCCTCGTCCGCAACCGGGCGAATCAGAGACAGACCGCGTGGCAACCTTAGT	459
QY	157	AlaGluAlaValThrArgGlyValGlyGlyMetSerThrHisTrpThrCysSerThrPro	176
DB	460	GCCAGGCGGTCAACCGCGTCTCGCGGCATGTCTACGCATGTGACGTGCGCGAGCGCG	519
QY	177	ArgIleHisProProMetGluSerLeuProGlyIleGlyArgProLysLeu---SerAsn	195
DB	520	CGCTTCGAGAAG-----CTGCAGCGCCGCTGCTCGTGAAGAAC	558
QY	196	AspProAlaGluAspLysGluTrpAsnGluLeuTyrSerGluAlaGluArgLeuIle	215
DB	559	GACTCCAGGCGGACGACCGCGAGTGGGACAGGCTCTACAAGAAGCGCGAGTCGACTTC	618
QY	216	GlyThrSerThrLysGluPheAspGluSerIleArgHisThrLeuValLeuArgSerLeu	235
DB	619	AGACCGGCACACCCAGTTCGCCAGTCGATCCGCCCAACCTCTGCTCAAGAAGCTG	678
QY	236	GlnAspAlaTyrLysAspArgGlnArgIlePheArgProLeuProLeuAlaCysHisArg	255
DB	679	CAGGAGGAGTACAAGGC---GTGCGCGACTTCCAGCAGATCCCGCTCGCGCGAGCGC	735
QY	256	LeuLysAsnAlaProGluTyrValGluTrpHisSerAlaGluAsnLeuPhe-----His	273
DB	736	-----CAGAGCCCGACGCTCGTCGAGTGGAGCTCGCGCACACCGCTGTCGATCTCGAG	789
QY	274	SerIleTyrAsnAspLysGlnLysLysLeuPheThrLeuLeuThrAsnHisArgCys	293
DB	790	AACCGCGCAACAAGCAGCGCGGAGGAGCGCTTCAACCTCTTCCCGCCCGTCGCGTGC	849
QY	294	ThrArgGluAlaLeuThrGlyGlyTyrGluLysLysIleGlyAlaAlaGluValArgAsn	313
DB	850	ACGAACGTGAGCGCGGATACACGCAACTCGGAGATGATAGGC---CTGATGTCCGCGAC	906
QY	314	LeuLeuAlaThrArgAsnProSerSerGlnLeuAspSerTyrIleMetAlaLysValTyr	333

RESULT 5

AAC87518

ID AAC87518 standard; DNA; 1995 BP.

XX

AC AAC87518;

XX

DB	907	CTCCACGGGGCAAGAGCATCAC-----ATCAAGGCCAAGGTGTAC	948
QY	334	ValLeuAlaSerGlyAlaIleGlyAsnProGlnIleLeuTyrAsnSerGlyPheSerGly	353
DB	949	ATCCTCACCGCGCGCGGTCCACAACGCGCAGCTCTCTCGCGCCCTCTGGATTTGGGCGAG	1008
QY	354	LeuGlnValThrProArgAsnAsp-----SerLeuIleProAsnLeu	367
DB	1009	CTG-----GGTGTCCCGACCCGCCAAGCCGCTCCGCTCTCTGTGCGCGTACCTG	1059
QY	368	GlyArgTyrIleThrGluGlnProMetAlaPheCysGlnIleValArgGlnGluPhe	387
DB	1060	GGGACCCACATCACCGAGCAGCGCTCTCTCTGCCAGACCGCTCATGAGACGAGGATC	1119
QY	388	ValAspSerValArgAspPro-----TyrGlyLeuPro-----	399
DB	1120	ATCAACAGTGTACCCCGGATATGACCATTTGCGGCAAGCCCGGCCACCCCGGACTATAGC	1179
QY	400	-----TriptLysGluAlaVal	405
DB	1180	GTCACTATATACCCCGGCAACCCGAACAACAAGCACCCGACTGGTGGAAACGAGAGGTG	1239
QY	406	AlaGlnHisIleAlaLysAsnProThrAspAlaLeuProIleProPheArgAspProGlu	425
DB	1240	AAGAAGCATGATGACCAACGAGGAGGACCGCTCCGATCCGTTCCGAGGACCCCTGAG	1299
QY	426	ProGlnValThrProPheThrGluGluHisProTrpHisThrGlnIleHisArgAsp	445
DB	1300	CCGAGGTCAACACGCTGTTTCAGGCAACGACCCCATGGCACACCCAGATTACCCGCGAC	1359
QY	446	AlaPheSerTyrGlyAlaValGlyProGluValAspSerArgValIleValAspLeuArg	465
DB	1360	GCCTTCAGTACCGCGCGCTGCAGCAGACATCGACTCGGGCTCATCTCGACTGCGCG	1419
QY	466	TrpPheGlyAlaThrAspProGluAlaAsnAsnLeuValPheGlnAsnAspValGln	485
DB	1420	TTCTTCGGACGACCGAGCCAGGCCAAGGAGGAGAAACAGTATGTTCTCGGACAAAGATCAG	1479
QY	486	AspGlyTyrSerMetProGlnProThrPheArgTyrArgProSerThrAlaSerAsnVal	505
DB	1480	GACGCGTACAACTCCGCGCAGCGGAGTTCGACTTCGCG-----TTCCCGGGGGCGCG	1533
QY	506	ArgAlaArgLysMetMetAlaAspMetCysGluValAlaSerAsnLeuGlyTyrLeu	525
DB	1534	GAAGCGGAGACATGATGACCGACATGTCGTCATGTCGCGAAGATCGTGGATTCTCTG	1593
QY	526	ProThrSerProGlnPheMetAspProGlyLeuAlaLeuHisLeuAlaGlyThrThr	545
DB	1594	CCTGGTCTCTACCCACAGTTTCATGGAGCCCGCTTGTCTCTGCACCTTGTGGGAGCGAC	1653
QY	546	ArgIleGlyPheAspLysAlaThr-----ThrValAlaAspAsnSerLeuValTyr	563
DB	1654	CCATGGGCTTCGACGAGGAGCGGCAAGTCTGCTGCGACACCGCTACCGCGTCTTC	1713
QY	564	AspPheAlaAsnLeuTyrValAlaGlyAsnGlyThrIleArgThrGlyPheGlyGluAsn	583
DB	1714	GCCTTCAAGAACCTCTCTCTCGCGCGCTCGCGGAAACATCCACCGCGTACCGCGGAAC	1773
QY	584	ProThrLeuThrSerMetCysHisAlaIleLysSerAlaArgSerIleIleAsnThrLeu	603
DB	1774	CCGAGCTCACCGCAATGTCGTTGCGATCAAGAGCTGCGAGTACATC-----AAG	1824
QY	604	LysGlyGlyThrAspGlyLysAsnThrGlyGluHisArgAsn	617
DB	1825	AAGAACTTCGAGCCGCGGACCCGCAACCCCGTGAAGACACCAAC	1866

QY 426 ProGlnValThrProPheThrGluGluHisProThrHisThrGlnIleHisArgAsp 445
 Db 1329 CCGCAGGTCCACCGCTGTTTCAGGCAACGCGACCCCTGCGACACCCAGATTTCACCGCGAC 1388
 QY 446 AlaPheSerTyrGlyAlaValGlyProGluValAspSerArgValIleValAspLeuArg 465
 Db 1389 GCCTTCAGTACGGCCCGTGCAGCAGCATCGACTCGGGCTCATCGTCGACTGGCGG 1448
 QY 466 TrpPheGlyAlaThrAspProGluAlaAsnAsnLeuValPheGlnAsnAspValGln 485
 Db 1449 TTCTTCGGACGACCGCAGCCAGGAGGAGAGCAAGCTATGTTCTCGCAGACATCAGC 1508
 QY 486 AspGlyTyrSerMetProGlnProThrPheArgTyrArgProSerThrAlaSerAsnVal 505
 Db 1509 GACGCTCAACACTCCGCGACGCGAGCTTCGACTTCGCGC-----TTCGCCGGGGCGCG 1562
 QY 506 ArgAlaArgLysMetMetAlaAspMetCysGluValAlaSerAsnLeuGlyGlyTyrLeu 525
 Db 1563 GAAGCGGAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1622
 QY 526 ProThrSerProGlnPheMetAspProGlyLeuAlaLeuHisLeuAlaGlyThrThr 545
 Db 1623 CCTGGTCTCTACCCACAGTTCATGGAGCCGCTTGTCTGCACCTTGTGGGACGAC 1682
 QY 546 ArgIleGlyPheAspLysAlaThr-----ThrValAlaAspAsnAsnSerLeuValTrp 563
 Db 1683 CGCATGGGTCTCGACGAGAGCGGACAAAGTGTGCTGCGACACCGACTCAGCGCTCTTC 1742
 QY 564 AspPheAlaAsnLeuTyrValAlaGlyAsnGlyThrIleArgThrGlyPheGlyGluAsn 583
 Db 1743 GCCTTCAAGAACTCTCTCTCGGCGCTCGGGAGCATCCCGACCGGTACCGCGCGGAAC 1802
 QY 584 ProThrLeuThrSerMetCysHisAlaIleLysSerAlaArgSerIleIleAsnThrLeu 603
 Db 1803 CCGACGCTCACCGCAATGCTGCTGCGATCAAGAGCTGCGAGTACATC-----AAG 1853
 QY 604 LysGlyGlyThrAspGlyLysAsnThrGlyGluHisArgAsn 617
 Db 1854 AAGAACTTCGAGCGCGAGCCGGAACCCCGTGAAGCACCACAAC 1895
 RESULT 6
 AAV83626
 ID AAV83626 standard; cDNA to mRNA; 1701 BP.
 XX
 AC AAV83626;
 XX
 DT 26-FEB-1999 (first entry)
 XX
 DE Nucleic acid encoding an antitumour protein.
 XX
 KW Antitumour protein; Tricholoma matsutake; ss.
 XX
 OS Tricholoma matsutake.
 XX
 FH Key Location/Qualifiers
 XX CDS 1..1701
 FT FT /tag= a
 FT FT /product= antitumour_protein
 XX
 PN JP1013876-A.
 XX
 PD 02-DEC-1998.
 XX
 PF 13-FEB-1998; 98JP-0031452.
 XX
 PR 13-FEB-1997; 97JP-0029275.
 XX
 PA (MOMO-) MOMOYA KK.
 PA (NORQ) NORINSUISANSHO SHOKUJIN SOGO.
 XX
 DR WPI; 1999-074153/07.
 DR P-PSDB; AAW87531.
 XX

PT An anti-tumour protein - prepared by culture of host cell
 PT transformed by vector containing base coding sequence
 XX
 PS Claim 3; Page 8-10; 15pp; Japanese.
 XX
 CC The present sequence encodes an antitumour protein, and is isolated
 CC from Tricholoma matsutake.
 XX
 SQ Sequence 1701 BP; 411 A; 463 C; 432 G; 395 T; 0 other;
 Alignment Scores:
 Pred. No.: 2,68e-86 Length: 1701
 Score: 1091.00 Matches: 248
 Percent Similarity: 57.04% Conservative: 84
 Best Local Similarity: 42.61% Mismatches: 184
 Query Match: 33.22% Indels: 66
 DB: 20 Gaps: 15
 US-09-856-327-2 (1-618) x AAV83626 (1-1701)
 QY 37 AspValPheIleAlaGlySerGlyProIleGlyAlaThrTyrAlaLysLeuCysValGlu 56
 Db 97 GATGTTTTCATGCTGCGCAGTGGTCCCATAGCTAGTACTAGCCGCCCATCATGTGAC 156
 QY 57 -----AlaGlyLeuArgValValMetValGluIleGlyAlaAlaAspSerPheTyrAla 74
 Db 157 AATACCTCAACTACAAAGGTTTACATGCGCGCAATAGGTTCTCAAGATAAC----- 207
 QY 75 ValAsnAlaGluGluGlyThrAlaValProTyrValProGlyTyrHisLysLysAsnGlu 94
 Db 208 -----CCTGTCATC-----GGGCGCCCATCACAGAACTCC 237
 QY 95 IleGluPheGlnLysAspIleAspArgPheValAsnValIleLysGlyAlaLeuGln 114
 Db 238 ATAAAGTTTCAAGAGACATTCACAAGTTTGAATATCATCAACGTCCTCCACGCG 297
 QY 115 ValSerValProValArgAsnGlnAsnValProThrLeuAspProGlyAlaTrpSerAla 134
 Db 298 ATTTCGATTTCCGCATCGGACACCTACACAGCCACTCTCGCTGTAGACGCTGG---GCG 354
 QY 135 ProProGlySerSerAla-----IleSerAsnGlyLysAsnProHisGlnArg 150
 Db 355 CCGCCCATCGATCTCGCCGAAGCCAGCTGCTGATATGGGACACATCCGAATCAGGAG 414
 QY 151 GluPheGluAsnLeuSerAlaGluAlaValThrArgGlyValGlyMetSerThrHis 170
 Db 415 GCGGCGCTGAACCTTCCCGGTAGCGCTGTCACTAGGACAGTCGGGGGATGCGACCCAC 474
 QY 171 TrpThrCysSerThrProArgIleHisProMetGluSerLeuProGlyIleGlyArg 190
 Db 475 TGGACTTGGCGTGTCTACTCCACATGAC----- 504
 QY 191 ProLysLeuSerAsnAspProAlaGluAspAspLysGluTrpAsnGluLeuTyrSerGlu 210
 Db 505 GAAGAGAGGTCACAAACCCAGTT---GACAGAGGAGTTTCGACGACATGTCGAACT 561
 QY 211 AlaGluArgLeuIleGlyThrSerThrLysGluPheAspGluSerIleArgHisThrLeu 230
 Db 562 GCTAAACATTCCTCAACGTTCCACGCGACAGTAGCAGGATCTATCCGTCAGATAGT 621
 QY 231 ValLeuArgSerLeuGlnAspAlaTyrLysAspArgGlnArgIlePheArgProLeuPro 250
 Db 622 GTCAAAGAGACTCTTCAGCAGACCCCTT---GATGCTCGCGGGGTGTGACCACTCTCCCG 678
 QY 251 LeuAlaCysHisArgLeuLysAsnAlaProGluTyrValGluTrpHisSerAlaGluAsn 270
 Db 679 CTGGGGGTGGAGCGCGTACGAGCAATCTATTATGTCACCTGACCGCGTCCCAT--- 735
 QY 271 LeuPheHisSerIleTyrAsnAspLysGlnLysLysLeuPheThrLeuThrAsn 290
 Db 736 -----ACGTCCTTGGTGATGCGGAAGAGTCCCGGATTCGCTTGTGTACAGAG 786
 QY 291 HisArgCysThrArgLeuAlaLeuThrGlyGlyTyrGluLysLysIleGlyAlaAlaGlu 310

```
Db 787 ACGAGAGTGACGAAGCTTATTGTCAGTGAACCAATCGACGAGGTGTGTCGCGTTG 846
Qy ValArgAsnLeuAlaThrArgAsnProSerGlnLeuAspSerTyrIleMetAla 330
Db 847 CTACGTAACCTG-----AATCAAGCAAC-----GATGAACCTGTGCTGGCC 888
Qy LysValTyrValLeuAlaSerGlyAlaIleGlyAsnProGlnIleLeuTyrAsnSerGly 350
Db 889 AAGATTTCGTCATAGTTGTGGAGCAGTCGCACACGCGCAAACTTTGTGGAACAGC--- 945
Qy PheSerGlyLeuGlnValThrProArgAsnAspSerLeuIleProAsnLeuGlyArgTyr 370
Db 946 -----AACATCCGCCCATAT-----GCGCTTGGTCGCTAC 975
Qy IleThrGluGlnProMetAlaPheCysGlnIleValLeuArgGlnGluPheValAspSer 390
Db 976 CTCACGACAGTCCATGACCTTTTGTGCAGATCGTTCTCAAGAGGGCATAGTCGATGCC 1035
Qy ValArgAspAspProTyrGlyLeuProTyrTrpLysGluAlaValAlaGlnHisIleAla 410
Db 1036 ATCGCTACTGACCTCGC-----TTCGCTGCGAAGGTTGAGGCGCACAGAAG 1083
Qy LysAsnProThrAspAlaLeuProIleProPheArgaspProGluProGlnValThrThr 430
Db 1084 AAGCACCCCGATGACGTGCTCCCATTCATTCACGAGCGCTGAACCTCAAGTGATGATT 1143
Qy ProPheThrGluGluHisProTyrHisThrGlnIleHisArgAspAlaPheSerTyrGly 450
Db 1144 CCGTACAGCTCGGACTCCCTTGGCAGTTCCTCAGTGCATCGCATTCCTCATATGGT 1203
Qy AlaValGlyProGluValAspSerArgValIleValAspLeuArgTyrPheGlyAlaThr 470
Db 1204 GATGTTGGACCAAGCGGACCGCGCTGTTGTCGTCGATCTGAGGTTTTTCGGCAAAATCA 1263
Qy AspProGluAlaAsnAsnLeuLeuValPhe-----Gln 481
Db 1264 GATATTGTCGAAGAAATCGAGTACTTTCGTCGCGCAACCCCTAAGCTACGCGAGTGGGAA 1323
Qy AsnAspValGlnAspGlyTyrSerMetProGlnProThrPheArgTyrArgProSerThr 501
Db 1324 CGCGGTGTACAGACACTATGGAATGCCAGCGGACATTCATGTCACGCGGACCAAC 1383
Qy AlaSerAsnValArgAlaArgLysMetMetAlaAspMetCysGluValAlaSerAsnLeu 521
Db 1384 GCCGATGGAGACCGTGACCGAGGATGATGAATGATATGACCAACGTCGCGAATGCTG 1443
Qy GlyGlyTyrLeuProThrSerProGlnPheMetAspProGlyLeuAlaLeuHisLeu 541
Db 1444 GGTGGGTACCTTCTGCTGCTCCTACCTCAATTTATGGCACCTGCTGCTACTGCACATC 1503
Qy AlaGlyThrThrArgIleGlyPheAspLysAlaThrThrValAlaAspAsnAsnSerLeu 561
Db 1504 ACGGAACTACTCGATCGGACAGATGATCAACTCTGTGTGATCCACATCAAG 1563
Qy ValTrpAspPheAlaAsnLeuTyrValAlaGlyAsnGlyThrIleArgThrGlyPheGly 581
Db 1564 GTTCATAACTTCAACAATCTGTGGTCGGCGGAATGGTGATCCAGATCGGACGTGCC 1623
Qy GluAsnProThrLeuThrSerMetCysHisAlaIleLysSerAlaArgSerIleIleAsn 601
Db 1624 TGCACCCGACTCGTACGAGCGTCGCTATCGCTCAAGGGTGTGTGAGGTGTGTATCAAT 1683
Qy ThrLeu 603
Db 1684 TACCTT 1689
```

RESULT 7

AAA71487

ID AAA71487 standard; cDNA; 1946 BP.

XX

AC AAA71487;

XX

```
DT 11-DEC-2000 (first entry)
XX T. matsutake pyranose oxidase cDNA.
DE Pyranose oxidase; matsutake mushroom; antibacterial; sugar metabolism;
KW diabetes mellitus; ss.
XX Tricholoma matsutake.
OS Location/Qualifiers
FH Key 101..1795
FT /*tag= a
FT /product= "pyranose oxidase"
PN JP2000175698-A.
XX 27-JUN-2000.
PD 16-DEC-1998; 98JP-0357423.
PF 16-DEC-1998; 98JP-0357423.
PR 16-DEC-1998; 98JP-0357423.
XX (NIBS ) JAPAN TOBACCO INC.
PA WPI; 2000-478485/42.
XX P-PSDB; AAB10457.
XX A reagent containing pyranose oxidase for the determination of pyranose
PS Disclosure; Page 14-17; 23pp; Japanese.
XX This invention describes a novel reagent containing pyranose oxidase for
the determination of pyranose which can be prepared from a fraction
precipitated from an aqueous extract of matsutake mushroom by ammonium
sulfate precipitation and has an antibacterial activity against at least
Pyricularia oryzae or Thanatephorus cucumeris and has a M.W. of ca.
210 kD by gel filtration and shows the presence of components of ca.
50 kD and ca. 15 kD by SDS-PAGE and maintains the above antibacterial
activity by being heated at 60 degrees C for 10 minutes in a neutral
aqueous solution and in which the above antibacterial activity is
inactivated by being heated at 80 degrees C for 10 minutes in a neutral
aqueous solution. The invention also describes a method for the
determination of pyranose in a sample in which the above reagent for the
determination of pyranose is reacted with pyranose in the sample and the
hydrogen peroxide formed is reacted with an enzyme to develop a color, a
method for diagnosing a disease accompanied by abnormality in sugar
metabolism in which the above reagent for the determination of pyranose
is reacted with pyranose in the sample and the hydrogen peroxide formed
is reacted with an enzyme to develop a color, and a kit for pyranose
analysis or the diagnosis of a disease accompanied by abnormality in
sugar metabolism. The pyranose oxidase can be used as a diagnostic agent
for diabetes mellitus. This sequence encodes the Tricholoma matsutake
pyranose oxidase protein which is described in the method of the
invention.
SQ Sequence 1946 BP; 484 A; 516 C; 475 G; 471 T; 0 other;
```

Alignment Scores:

```
Pred. No.: 4,48e-82 Length: 1946
Score: 1044.00 Matches: 241
Percent Similarity: 55.91% Conservative: 90
Best Local Similarity: 40.71% Mismatches: 187
Query Match: 31.79% Indels: 74
Db: 21 Gaps: 17
```

US-09-856-327-2 (1-618) x AAA71487 (1-1946)

Qy 27 IleHisGluThrTyrGlyAsnAspGlyValAspValPheIleAlaGlySerGlyProfile 46

Db 185 GTACATTACACT-----GATGTTTTTCATGCTGGCAGTGGTCCCAT 226

Qy 47 GlyAlaThrTyrAlaLysLeuCysValGlu-----AlaGlyLeuArgValMetVal 64

```
Db 227 GCGTGTACTACGCGCCGACATCATTTGACAATACCTCAACTACAAAGGCTCTACATGGCC 286
QY 65 GluIleGlyAlaAlaAspSerPheTyrAlaValAsnAlaGluGluGlyThrAlaValPro 84
Db 287 GAATAGGTTCTCAGATAAC-----CCT 310
QY 85 TyrValProGlyTyrHisIstysAsnGluIleGluPheGlnLysAspIleAspArgPhe 104
Db 311 GTCATC---GGAGCCCATCACAGGAACCTCCATAAAGTTTCAGAAAGACACTCACAAGTTT 367
QY 105 ValAsnValIleGlyAlaLeuGlnGlnValSerValProValArgAsnGlnAsnVal 124
Db 368 GTGAATATCATCAACGGTGGCTCCAGGCCATTTCGATTTCGCCATCGGACACCTACCAG 427
QY 125 ProThrLeuAspProGlyAlaTrpSerAlaProProGlySerSerAla----- 140
Db 428 CCCACTCTCGTGTAGCAGCGTGG---CGCCGCCCATCGATCGCTCGGAAGGCCAGCTC 484
QY 141 IleSerAsnGlyLysAsnProHisGlnArgGluPheGluAsnLeuSerAlaGluAlaVal 160
Db 485 GTGATTATGGGACACAATCCGAATCAGGAGGCCGCCCTGAACCTTCCCGGTAGCGCTGTC 544
QY 161 ThrArgGlyValGlyClyMetSerThrHisTrpThrCysSerThrProArgIleHisPro 180
Db 545 ACGAGGACAGTCGGGGGAATGGCGACCTGGACTTGGCGTGTCTACTCCACATGAC 604
QY 181 ProMetGluSerLeuProGlyIleGlyArgProLysLeuSerAsnAspProAlaGluAsp 200
Db 605 -----GAAGAGAGGGTCAACAACCCAGTT---GAC 631
QY 201 AspLysGluTrpAsnGluLeuTyrSerGluAlaGluArgLeuIleGlyThrSerThrLys 220
Db 632 AAGCAGGAGTTTCGACGCACTGCTCGAACGTGTAAACATTCCTCAACGTTTCACAGCGAC 691
QY 221 GluPheAspGluSerIleArgHisThrLeuValLeuArgSerLeuGlnAspAlaTyrLys 240
Db 692 CAGTATGACGATTCATCCGTGACAGTAGTTGTCAAGAGAGACCCCTTCACAGACCCCTT--- 748
QY 241 AspArgGlnArgIlePheArgProLeuProLeuAlaCysHisArgLeuLysAsnAlaPro 260
Db 749 GATGCGTCGCGGGTGTACCACTCTCCGCTGGGGTGGAGCGCGCGGACGCAATCCT 808
QY 261 GluTyrValGluTrpHisSerAlaGluAsnLeuPheHisSerIleTyrAsnAspLys 280
Db 809 ATTATGTACCTGGACCGGTGCCGAT-----ACCGTCTTGGTGTATGTGCGG 856
QY 281 GlnLysLysLeuPheThrLeuThrAsnHisArgCysThrArgLeuAlaLeuThrGly 300
Db 857 AAGATCCCGGATTCGTTTGGTTACAGAGAGAGAGTGCAGCAAGTTTATTGTCAGTGAA 916
QY 301 GlyTyrGluLysLysIleGlyAlaAlaGluValArgAsnLeuLeuAlaThrArgAsnPro 320
Db 917 ACCAATCCGACGAGGTGTGCTGCGTGTGTACGTAACCTTG-----AATACA 964
QY 321 SerSerGlnLeuAspSerTyrIleMetAlaLysValTyrValLeuAlaSerGlyAlaIle 340
Db 965 AGCAAC-----GATGAATCTGCTGGCCAGAGCTTTCGTATAGTTGTGGAGCAGTC 1018
QY 341 GlyAsnProGlnIleLeuTyrAsnSerGlyPheSerGlyLeuGlnValThrProArgAsn 360
Db 1019 TGCACACGCAATCCTGTGAACAGC-----AACATCCGCCCAT--- 1060
QY 361 AspSerLeuIleProAsnLeuGlyArgTyrIleThrGlnProMetAlaPheCysGln 380
Db 1061 -----GCCCTTGGTGGCTACCTCAGCAACAGCTCCATGACTCTTTTGTCTAG 1105
QY 381 IleValLeuArgGlnGluPheValAspSerValArgAspProTyrGlyLeuProTrp 400
Db 1106 ATTGTTCTCAAGAGGAGCATAGTCGATTCCATCGCTACTGACCCCTCGC----- 1153
QY 401 TrpLysGluAlaValAlaGlnHisIleAlaLysAsnProThrAspAlaLeuProIlePro 420
Db 420 -----
```

```
1154 TTCGTGTGGAGGTTGAGGCGCACAGAAGAACGCCGATGACGTGTGCGCATTCOA 1213
QY 421 PheArgAspProGluProGlnValThrThrProPheThrGluGluHisProTrpHisThr 440
Db 1214 TTCACAGAGCCTGAACCTCAAGTATGATTCGTACACAGTCGCGATTCCTTGGCATGTT 1273
QY 441 GlnIleHisArgAspAlaPheSerTyrGlyAlaValGlyProGluValAspSerArgVal 460
Db 1274 CAGGTCCATCGC-----TATGCATTGGTGTATGTTGGACCAAGCCGACCGCGTGT 1327
QY 461 IleValAspLeuArgTrpPheGlyAlaThrAspProGluAlaAsnAsnLeuLeuValPhe 480
Db 1328 GTCGTGATCTGAGGTTTTTCGGCAATCAGATATTGTGAAAGAAATCGAGTGAATTC 1387
QY 481 -----GlnAsnAspValGlnAspGlyTyrSerMetPro 491
Db 1388 GGTCCGAACCCTAAGCTACCGGAGCGGGTGTACAGACACITATATGAATGCCA 1447
QY 492 GlnProThrPheArgTyrArgProSerThrAlaSerAsnValArgAlaArgLysMetMet 511
Db 1448 CAGCCGACATTCATGTCACGCGGACCAACGCCGATGGAGACCGTGACCAGAGGATGATG 1507
QY 512 AlaAspMetCysGluValAlaSerAsnLeuGlyGlyTyrLeuProThrSerProGln 531
Db 1508 AATGATATGACCAACGTCGCAACACTACTGGCGGGTACTTCTGCTCTACCCCTCAA 1567
QY 532 PheMetAspProGlyLeuAlaLeuHisLeuAlaGlyThrArgIleGlyPheAspLys 551
Db 1568 TTTATGGCACCTGGTCTCGCACAGCACATCAGCGGAACACTCTCGGATCGGACAGATGAT 1627
QY 552 AlaThrThrValAlaAspAsnAsnSerLeuValTrpAspPheAlaAsnLeuTyrValAla 571
Db 1628 CAAACTTCTGTGTGCTGATCCGACATCAAAGGTTTCACTTCCACAATCTGTGGTCTGCGC 1687
QY 572 GlyAsnGlyThrIleArgThrGlyPheGlyGluAsnProThrLeuThrSerMetCysHis 591
Db 1688 GGAATGGGTGCATTCCAGATCGCACTGCCCTCAACCCGACTCTGACGAGCGTACGAT 1747
QY 592 AlaIleLysSerAlaArgSerIleIleAsnThrLeu 603
Db 1748 GCGCTTAAGGTGCTGAGGCTGTAGTCATACCTT 1783

RESULT 8
ID AAA07403 standard; DNA; 1946 BP.
XX
AC AAA07403;
XX
DT 07-JUL-2000 (first entry)
XX
DE Trichoderma derived antifungal protein coding sequence.
XX
KW Antifungal protein; Pyricularia oryzae; Rhizoctonia solani; mushroom;
KW growth inhibitor; plant pathogenic fungi; antibacterial agent;
KW N-terminal fragment; ss.
XX
Trichoderma matsutake.
XX
WO200014242-A1.
XX
PD 16-MAR-2000.
XX
PF 19-AUG-1999; 99WO-JP04441.
XX
PR 08-SEP-1999; 98JP-0270606.
XX
PA (NTSB ) JAPAN TOBACCO INC.
PA (NORQ ) SOC TECHNO-INNOVATION AGRIC FORESTY & FI.
XX
PI Takakura Y, Kuwata S, Ohta S;
XX
DR WPI; 2000-256990/22.
DR P-PSDB; AAY81952.
```

XX Mushroom-derived antibacterial protein against plant pathogenic fungi
 PT of rice, with activity and thermal stability, obtainable cheaply on
 PT large scale, useful in agriculture -
 XX
 PS Claim 13; Page 41-45; 52pp; Japanese.
 XX
 CC This sequence encodes the *Trichoderma matsutake* antibacterial protein of
 CC the invention. The protein has activity against at least *Pycnularia*
 CC *oryzae* and *Rhizoctonia solani*, and is obtained from a fraction of an
 CC aqueous extract of a mushroom precipitated by the ammonium sulphate
 CC precipitation method. The protein has a molecular weight of about 210 kD
 CC as determined by the gel filtration method, includes components of about
 CC 15 kD and 50 kD in SDS-PAGE, and is stable to heating in an aqueous
 CC neutral solution at 60 degrees C for 10 minutes but with loss of
 CC antibacterial activity after heating in the solution at 80 degrees C for
 CC 10 minutes. The protein is used for inhibiting the growth of plant
 CC pathogenic fungi e.g. *Pycnularia oryzae* and *Rhizoctonia solani*.
 CC It is useful in treating rice plants, and is applicable in agriculture as
 CC an antibacterial agent. The protein has activity at relatively low
 CC concentrations, and can be produced at low cost on large scale.
 XX
 SQ Sequence 1946 BP; 484 A; 516 C; 475 G; 471 T; 0 other;

Alignment Scores:

Pred. No.:	4.48e-82	Length:	1946
Score:	1044.00	Matches:	241
Percent Similarity:	55.91%	Conservative:	90
Best Local Similarity:	40.71%	Mismatches:	187
Query Match:	31.79%	Indels:	74
DB:	21	Gaps:	17

US-09-856-327-2 (1-618) x AAA07403 (1-1946)

Qy	27	IleHisGluThrTyrGlyAsnAspGlyValAspValPheIleAlaGlySerGlyProIle	46
Db	185	GTACATTACACT-----GATGTTTTCAATGTGTCAGTGGTCCCAT	226
Qy	47	GlyAlaThrTyrAlaLysLeuCysValGlu-----AlaGlyLeuArgValValMetVal	64
Db	227	GCCTGTACTACGGCCGCCCATCATCTGACAAATACCTCACTACAAGGTCATACATGCC	286
Qy	65	GlutIleGlyAlaAlaAspSerPheTyrAlaValAsnAlaGluGlyThrAlaValPro	84
Db	287	GAATAGAGTTCTCAAGATAAC-----CCT	310
Qy	85	TyrValProGlyTyrHisLysLysAsnGluIleGluPheGlnLysAspIleAspArgPhe	104
Db	311	GTATC---GGAGCCCATACAGGAACTCCATAAGTTTCAGAAAGACACTGCAAGTTT	367
Qy	105	ValAsnValIleLysGlyAlaLeuGlnGlnValSerValProValArgAsnGlnAsnVal	124
Db	368	GTGAATATCATCAACGGTGCCTCCAGCCCATTTTCGATTCGCCATCGGACACCTACCAG	427
Qy	125	ProThrLeuAspProGlyAlaTrpSerAlaProGlySerSerAla-----	140
Db	428	CCCACTCTCGGTAGCAGCGTGG---GCGCGCCCATCGATCTCTCGCAGAGGCCAGCTC	484
Qy	141	IleSerAsnGlyLysAsnProHisGlnArgGluPheGluAsnLeuSerAlaGluAlaVal	160
Db	485	GTGATTATGGGACAAATCCGAATCAGGAGCCGCCCTGAACTTCCCGTACGCGTGTGTC	544
Qy	161	ThrArgGlyValGlyGlyMetSerThrHisTrpThrCysSerThrProArgIleHisPro	180
Db	545	ACGAGGACAGTCGGGGGAATGGCCACCCACTTGACTTGGCGTCTCTACTCCACATGAC	604
Qy	181	ProMetGluSerLeuProGlyIleGlyArgProLysLeuSerAsnAspProAlaGluAsp	200
Db	605	-----GAAGAGAGGGTCAACACCCAGTT---GAC	631
Qy	201	AspLysGluTrpAsnGluLeuTyrSerGluAlaGluArgLeuIleGlyThrSerThrLys	220
Db	632	AAGCAGGAGTTTCGACGCACTGCTCGAAACGTGCTAAACATTGCTCAACAGTTTCACAGGCAC	691

Qy	221	GluPheAspGluSerIleArgHisThrLeuValLeuArgSerLeuGlnAspAlaTyrLys	240
Db	692	CAGTATGACGATTCTATCCGTTCAGATAGTTCTCAAGAGACCCCTTCACGACACCTT---	748
Qy	241	AspArgGlnArgIlePheArgProLeuAlaCysHisArgLeuLysAsnAlaPro	260
Db	749	GATCGTCGCGGGTGTGACCACTCTCCCGTGGGGTGGAGCGCGCACGACAAATCCT	808
Qy	261	GluTyrValGluTrpHisSerAlaGluAsnLeuPheHisSerIleTyrAsnAspLys	280
Db	809	ATTTATGTCACTGGACCGGTGCCGAT-----ACCGTCTTGGTGTGATGTGCGG	856
Qy	281	GlnLysLysLeuPheThrLeuLeuThrAsnHisArgCysThrArgLeuAlaLeuThrGly	300
Db	857	AAGAGTCCCGCATTCGTTTGGTTACAGAGAGAGAGTACGAAGTTTATGTCAGTGAA	916
Qy	301	GlyTyrGluLysLysIleGlyAlaGluValArgAsnLeuLeuAlaThrArgAsnPro	320
Db	917	ACCAATCCGACGCGAGTGTGCTCGCTGCTAGCTAACTTG-----AATACA	964
Qy	321	SerSerGlnLeuAspSerTyrIleMetAlaLysValTyrValLeuAlaSerGlyAlaIle	340
Db	965	AGCAAC-----GATGAACCTGTGTCGCCCCAGAGTTTCGTCAATAGCTTGTGGACGATC	1018
Qy	341	GlyAsnProGlnIleLeuTyrAsnSerGlyPheSerGlyLeuGlnValThrProArgAsn	360
Db	1019	TGCACACGCGAAATCCTGTGGAACAGC-----AACATCCGCCCATCAT---	1060
Qy	361	AspSerLeuIleProAsnLeuGlyArgTyrIleThrGluGlnProMetAlaPheCysGln	380
Db	1061	-----GCGCTTGGTGGCTACCTCAGCAACAGTCCATGACTTTTGTGTCAG	1105
Qy	381	IleValLeuArgGlnGluPheValAspSerValArgAspAspProTyrGlyLeuProIrp	400
Db	1106	ATTGTTCTCAAGAGAGCATAGTCGATTCCATCGCTACTGACCCCTCGC-----	1153
Qy	401	TrpLysGluAlaValAlaGlnHisIleAlaLysAsnProThrAspAlaLeuProIlePro	420
Db	1154	TTCGCTCGAAGTTGAGCGGCACAGAAGAAGACCCCGATGAGTGTGCGGATTCCA	1213
Qy	421	PheArgAspProGluProGlnValThrThrProPheThrGluGluHisProThrIleThr	440
Db	1214	TTCACGACGCTGAACCTCAAGTATGATGATTCGTCACAGCTGCGACTTCCCTTGGCATGT	1273
Qy	441	GlnIleHisArgAspAlaPheSerTyrGlyAlaValGlyProGluValAspSerArgVal	460
Db	1274	CAGTCCATCCG-----TATGCAATTTGGTGTGATGTTGGACCCCAAGGCCCGCGGTGT	1327
Qy	461	IleValAspLeuArgTrpPheGlyAlaThrAspProGluAlaAsnAsnLeuValPhe	480
Db	1328	GTCTGTCGATCTGAGGTTTTTCGCGCAAAATCAGATATTTGTCGAAGAAAATCGAGTGATTC	1387
Qy	481	-----GlnAsnAspValGlnAspValGlnAspGlyTyrSerMetPro	491
Db	1388	GGTCCGACACCTAAAGTACGCGACTGCGGAAGCGGTGTACAGACACTTATGGAATGCCA	1447
Qy	492	GlnProThrPheArgTyrArgProSerThrAlaSerAsnValArgAlaArgLysMetMet	511
Db	1448	CAGCGCATTTCCATGTCAGCGGACCAACCGCATGGAGCCGTGACGACGAGGATGATG	1507
Qy	512	AlaAspMetCysGluValAlaSerAsnLeuGlyGlyTyrLeuProThrSerProProGln	531
Db	1508	AATGATATGACCAACGTCGCGAACAATACTGGCGGGGTACCTTCTGGCTCTCTACCTCAA	1567
Qy	532	PheMetAspProGlyLeuAlaLeuHisLeuAlaGlyThrArgIleGlyPheAspLys	551
Db	1568	TTTATGGCACCTGCTCGCACAGCATACCGGAACTACTCGGATCGGACAGATGAT	1627
Qy	552	AlaThrThrValAlaAspAsnAsnSerLeuValTrpAspPheAlaAsnLeuTyrValAla	571
Db	1628	CAAACTTCTGTTGTGATCCGACATCAAAAGGTTTCAATCTTCGAAATCTGTGGGTGCGG	1687

QY 572 GlyAsnGlyThrIleArgThrGlyPheGlyGluAsnProThrLeuThrSerMetCysHis 591
 DB 1688 GGGATGGGTGCATTCAGATCGGACTCGCTGCAACCGAGCTCGTACGAGCGTCGCGTAT 1747

QY 592 AlalIleLysSerAlaArgSerIleIleAsnThrLeu 603
 DB 1748 GCGCTTAAGGGTGTGAGGCTGTAGTACGTACCTT 1783

RESULT 9
 ID ABK52367 standard; DNA; 2467 BP.
 XX ABK52367;
 XX
 XX 27-AUG-2002 (first entry)
 XX
 XX DNA encoding glucose dehydrogenase associated proteins.
 XX Glucose dehydrogenase; electrode; glucose sensor; glucose-assay;
 KW substrate specificity; heat stability; gene; ds.
 XX Burkholderia cepacia.
 XX
 XX Key Location/Qualifiers
 FH CDS 258..764
 FT /*tag= a
 FT /product= "Glucose dehydrogenase associated protein #1"
 CDS 764..2383
 FT /*tag= b
 FT /product= "Glucose dehydrogenase associated protein #2"
 CDS 2386..2466
 FT /*tag= c
 FT /product= "Glucose dehydrogenase associated protein #3"
 XX
 XX WO200236779-A1.
 PN
 XX 10-MAY-2002.
 PD
 XX 31-OCT-2001; 2001WO-JP09556.
 XX PF
 XX 31-OCT-2000; 2000JP-0332085.
 PR
 XX 24-NOV-2000; 2000JP-0357102.
 PR
 XX 12-SEP-2001; 2001JP-0276832.
 XX
 XX (SODE/) SODE K.
 PA
 PI Sode K;
 XX
 XX WPI; 2002-463413/49.
 DR
 XX P-PSDB; AAU97825, AAU97826, AAU97827.
 XX
 XX Production of Burkholderia glucose dehydrogenase for use in glucose
 PT sensor electrodes and glucose-assay kits in medicine, science and
 PT industry -
 XX
 XX Claim 17; Page 51-55; 61pp; Japanese.
 PS
 XX The invention describes a method of producing a glucose dehydrogenase
 CC comprising culturing a Burkholderia microorganism and collecting the
 CC product from the medium and/or the microbial cells. Glucose dehydrogenase
 CC is useful in electrodes of glucose sensors and glucose-assay kits for
 CC medicine, science and industry. Glucose dehydrogenase is economically
 CC produced with high substrate specificity and improved heat stability to
 CC provide long-term accuracy. This sequence encodes 3 glucose dehydrogenase
 CC associated protein described in the invention.
 XX
 XX Sequence 2467 BP; 497 A; 792 C; 769 G; 409 T; 0 other;

Alignment Scores:
 Pred. No.: 2,61e-08 Length: 2467
 Score: 208.50 Matches: 131
 Percent Similarity: 34.05% Conservative: 91
 Best Local Similarity: 20.09% Mismatches: 221

Query Match: 6.35% Indels: 209
 DB: 24 Gaps: 29

US-09-856-327-2 (1-618) x ABK52367 (1-2467)

QY 37 AspValPheIleAlaGlySerGlyProIleGlyAlaThrTyrAlaLysLeuCysValGlu 56
 DB 791 GACGTGCTGCTGGTGGATCGGTGCGGGCGGATCGTCGCGCATCAGCTCGCATG 850

QY 57 AlaGlyLeuArgValValMetValGluIleGlyAlaAlaAspSerPheTyrAlaValAsn 76
 DB 851 GCGGCAAGCGGTGATCTCTCGAAGCGGCG----- 883

QY 77 AlaGluGluGlyThrAlaValProGlyTyrHisLysLysAsnGluIleGlu 96
 DB 884 -----CCGCGCATCGCGCGCTG-----GAAATC--- 907

QY 97 PheGlnLysAspIleAspArgPheValAsnValIleLysGlyAlaLeuGlnValSer 116
 DB 908 -----GTCGAGCGCTTCGCAAT----- 925

QY 117 ValProValArgAsnGlnAsnValProThrLeuAspProGlyAlaTrpSer----- 133
 DB 926 CAGCCCGACAAAGATGGACTTCATGCGCGCTACCGTCGAGCCCTGGGCGCGCATCC 985

QY 134 -----AlaProProGlySerSerAlaIleSerAsnGlyLysAsnProHisGlnArgGlu 151
 DB 986 GAGTACGCGCGCGGACGACTACTGATCTCGAAGGCGGAGCACAACTTCAACTCGCAG 1045

QY 152 PheGluAsnLeuSerAlaGluAlaValThrArgGlyValGlyGlyMetSerThrHisTrp 171
 DB 1046 TAC-----ATCGCGCGGTGGGCGGACGACGCTGGCAGCTGG 1081

QY 172 ThrCysSerThrProArgIleHisPro-----PrometGluSerLeuProGlyIle 188
 DB 1082 GCCGCGTGGCGGTGGCGCTTCATCCGAACGACTTCAAGATGAAGCGGTGTACGCGT 1141

QY 189 GlyArgProLysLeuSerAsnAspProAlaGlu---AspAspLysGluTrpAsnGluLeu 207
 DB 1142 GGCGCG-----GACTGGCGGATCCAGTACGACGATCTCGAG-----CCGTAC 1183

QY 208 TyrSerGluAlaGluArgLeuIleGlyThrSerThrLysGluPheAspGluSerIleArg 227
 DB 1184 TATCAGCGCGGAGGAGAGCTCGCGTGTGGGCGCGGCGCGCGAGAA----- 1234

QY 228 HisThrLeuValLeuArgSerLeuGlnAspAlaTyrLysAspArgGlnArgIlePhe--- 246
 DB 1235 -----GATCTGTACTCGCGCGCAAGCAGCGCGTATCCG 1267

QY 247 ---ArgProLeuProLeuAlaCysHisArg-----LeuLysAsnAlaProGluTyrVal 263
 DB 1268 ATCGCGCGCGTGGCGTTCGTTCAACGAGCAGACCATCAAGACGCG----- 1315

QY 264 GluTrpHisSerAlaGluAsnLeuPheHisSerIle-----TyrAsnAsp 278
 DB 1316 ---CTGAACAACACTACGATCCGAAAGTTCCATGCTCGTGACCGCGCGCGCAACAGC 1372

QY 279 AspLysGlnLysLysLeuPheThrLeuLeuThrAsnHisArgCysThrArgLeuAlaLeu 298
 DB 1373 CGCGCGTACGAGCGCGCGCGCGTGTGTCGGCAACAACAACTGCGCATCGCGTCCGG 1432

QY 299 ThrGlyGlyTyr-----GluLysLysIleGlyAlaAla 309
 DB 1433 ATCGCGCGGATGTACAACGGCATCGTGACGTCGAGAGGCGGACGCGCGCGCGCAAG 1492

QY 310 GluValArgAsnLeuLeuAlaThrArg-----AsnProSerSerGlnLeu----- 324
 DB 1493 CTGATCGAAGACCGCGGTCTCTACAACTCGAGACGGCGCGGCAACGCGCATCTCGCG 1552

QY 325 -----AspSerTyrIleMetAlaLysValTyrValLeu 335
 DB 1553 GCGCTCTACAAGGACACAGCGCGCGCGCGCATCGGTGCGTGAAGGCAAGTATTTCGTGCTC 1612


```
Qy 336 AlaSerGlyAlaIleGlyAsnProGlnIleLeuTyrAsnSerGlyPheSerGlyLeuGln 355
Db 1613 GCCGCGACGCGATCGAGACCGCGAAGATCTG-----
Qy 356 ValThrProArgAsnAspSerLeuIleProAsnLeuGlyArgTyrIleThrGluGlnPro 375
Db 1645 -----
Qy 376 MetAlaPheCysGlnIleValLeuArgGlnGluPheValAspSerValArgAspPro 395
Db 1646 -----CTGATGTCGCGGACCGCGATTTCCCG 1672
Qy 396 TyrGlyLeuProThrTrpTrpLysGluAlaValAlaGlnHisIleAlaLysAsnProThrAsp 415
Db 1673 AACGGTCGCGAAGACGTGGACATCGGACGCGCAACCTGATGGACCATCCGCGGAC 1732
Qy 416 AlaLeuProIle-----ProPheArgAspProGluProGlnVal 428
Db 1733 GCGGTGTCGTTCTATGCGAGGAGAGAGTGTGGCGCGCGCGCGCGAGAGATGACG 1792
Qy 429 ThrThrProPheThrGluGluHisProThrHisThr-----GlnIleHis 443
Db 1793 TCGGTGTCGTTCTCCGCGAGCGTCCGTTCGCGCGACCGAGCGCGAAGATCCAC 1852
Qy 444 -----ArgAspAlaPheSerTyrGlyAlaVal 452
Db 1853 CTGTGCAACCTGTGCGCATCGCAGGAGACGAGAGATCTTCAAGGCGCGAAGCTG 1912
Qy 453 -----GlyProGluValAspSerArgVal-----IleValAspLeu 464
Db 1913 ATGAAGCCGAGAGCTCGAGCGCGAGATCGCGACCGTTCGCGACCGTACGTCAGTTC 1972
Qy 465 ArgTrpPheGlyAlaThrAspProGluAlaAsnLeuLeuValPheGlnAsnAspVal 484
Db 1973 GACTGCTTCCAGAAATCCCTCCCAACCCGAGAACCGCATCGTCCGAGCAACCGCG 2032
Qy 485 GlnAspGlyTyrSerMetProGlnProThrPheArgTyrArgProSerThrAlaSerAsn 504
Db 2033 ACCGATCGATCGCGCATTCGCGCGCGAGATCAGTAT-----GCGATCGAC 2080
Qy 505 ValArgAlaArgLysMetMetAlaAspMetCysGluVal-----AlaSerAsn 520
Db 2081 GACTAGTGAAGCGCGCGCGCATACGCGCGAGGTCTACGCGACCGCGCGAAGGTG 2140
Qy 521 LeuGlyGly-----TyrLeuProThrSerProGlnPhe 532
Db 2141 CTCGCGCGCAGGACGTGCTTCAACGACCAATTCGCGCGCAACAT-----2188
Qy 533 MetAspProGlyLeuAlaLeuHisLeuAlaGlyThrThrArgIleGlyPheAspLysAla 552
Db 2189 -----CACATCAGCGGCTCGACGATCATGGCGCGATCGCGCG 2227
Qy 553 ThrThrValAlaAspAsnAsnSerLeuValTrpAspPheAlaAsnLeuTyrValAlaGly 572
Db 2228 GACTCCGTGTCGACAGGACGTGCGGACGCTTCGACCATCGGAACCTGTTCATTTCCAGC 2287
Qy 573 AsnGlyThrIleArgThrGlyPheGlyGluAsnProThrLeuThrSerMetCysHisAla 592
Db 2288 AGCGCGAGATGCGCGCGCGTACCGTACCGTAAACGTGACGTGACGTGCGCGCGCTCGCG 2347
Qy 593 IleLysSerAlaArgSerIleIleAsnThrLeuLys 604
Db 2348 CTGCGGATGTCG-----GACACGCTGAAG 2371
RESULT 10
ABL52917
ID ABL52917 standard; DNA; 4290 BP.
XX
XX ABL52917;
XX AC
XX XX
DT 16-JUL-2002 (first entry)
XX
XX 2-keto-D-gluconate dehydrogenase coding sequence.
XX DE
```

Cell membrane bound; 2-keto-D-gluconate dehydrogenase; enzyme;
2,5-diketo-D-gluconate; gene; ds.

Unidentified.

Key Location/Qualifiers
CDS 337..3991
FT /*tag= a
FT /product= "2-keto-D-gluconate dehydrogenase"
FT /note= "Contains one intron"
FT 337..900
FT /*tag= b
FT /number= 1
FT 901..926
FT /*tag= c
FT /cons_splice= (5'site:NO,3'site:NO)
FT /number= 1
FT 927..3998
FT /*tag= d
FT /transl_except= (pos:2580..2581, aa:Met)
FT /number= 2

KR2000019366--A.
06-APR-2000.
10-SEP-1998; 98KR-0037413.
10-SEP-1998; 98KR-0037413.
(KOAD) KOREA ADV INST SCI & TECHNOLOGY.
Shin YC, Bahn JG, Yeom DY;
WPI; 2001-088033/10..
P-PSDB; AAM48440.
Base sequence of novel cell membrane-bound 2-keto-D-gluconate
dehydrogenase gene and method for conversion into
2,5-diketo-D-gluconate - NoAbstract
Claim 2; Page 6-8; 16pp; Korean.
The present invention relates to a novel cell membrane-bound
2-keto-D-gluconate dehydrogenase and a method for conversion into
2,5-diketo-D-gluconate. The present sequence is the coding sequence for
the 2-keto-D-gluconate dehydrogenase.
Sequence 4290 BP; 992 A; 1177 C; 1226 G; 895 T; 0 other;

Alignment Scores:
Pred. No.: 7.62e-08 Length: 4290
Score: 207.00 Matches: 141
Percent Similarity: 35.01% Conservative: 89
Best Local Similarity: 21.46% Mismatches: 251
Query Match: 6.30% Indels: 179
DB: 23 Gaps: 27

US-09-856-327-2 (1-618) x ABL52917 (1-4290)

Qy 21 GlnIleProLysAsnAlaIleHisGluThrTyrGlyAsnAspGlyValAspValPheIle 40
Db 921 AAAATCATGAAAAAACCCAGTGTCTTACTCGCGAAGCGATGCTTCGCGAGATATTGTGATC 980
Qy 41 AlaGlySerGlyProIleGlyAlaThrTyrAlaLysLeuCysValGluAlaGlyLeuArg 60
Db 981 GTGGGTTCCGGTATTGTTGGCGGTATGATGCCCAACCACTGGTTCAGTCAGGATATTCG 1040
Qy 61 ValValMetValGluIleGlyAlaAlaAspSerPheTyrAlaValAlaGluGluGly 80
Db 1041 GTACTGTGTCGGAAGCGCGCTG-----1064

QY 81 ThrAlaValProTyrValProGlyTyrHisLysLysAsnGluIleGluPheGlnLysAsp 100
 Db 1065 -----CGTATGATCGCGCGCGCC 1085
 QY 101 IleAspArgPheValAsnValIleLysGlyAlaLeuGlnGlnValSerValProValArg 120
 Db 1086 GTGGAAACTGGCGCAAT-----ATGCGGTTCGCA 1115
 QY 121 AsnGln-----AsnValProThrLeuAspProGlyAlaTrpSerAlaPro----- 135
 Db 1116 AACCGCGCAGGCTCCGACTTTCAGGATATATCTCAGTCAAAATTTGCTCCGCGCAGC 1175
 QY 136 -----ProGlySerSerAlaIleSerAsnGlyLysAsnProHisGlnArgLuphe 152
 Db 1176 CTCATATTTCCCGCTAACAACTATGTC---AACGTGACGGGTCCCAATGCTGACAGTTT 1232
 QY 153 GluAsnLeuSerAlaGluAlaValThrArgGlyValGlyGlyMetSerThrHisTrpThr 172
 Db 1233 CAA-----CAAGGCTACCTGGCGCAGGTAGCGGCAGCAGCTGGCAGTGGCA 1280
 QY 173 CysSerThrProArgIleHisPro-----MetGluSerLeuProGlyIleGly 189
 Db 1281 GCCTCTGCTGGCAGATCATCCAGCGATTTTGTGTCAGTCCGAATACGCGTGGC 1340
 QY 190 ArgPro---LysLeuSerAsnAspProAlaGluAspAspLysGluTrpAsnGluLeuTyr 208
 Db 1341 CCGCAGTGGCGCATAGGTATGACGAGTGGAA-----CCCTGG-----TAT 1382
 QY 209 SerGluAlaGluArgLeuIleGlyThrSerThrLysGluPheAspGluSerIleArgHis 228
 Db 1383 TCAGAGCTGAAACGAGATTGGCTGCAGGGCCA-----AAGCATCCCGCCAGACAA 1436
 QY 229 ThrLeuValLeuArgSerLeuGlnAspAlaTyrLysAspArgGlnArgIlePheArgPro 248
 Db 1437 TCACCCACGCGCGCAGC-----GATATGGTGCCTTTTCCCGCAGCGC 1493
 QY 249 LeuProLeuAlaCysHisArgLeuLysAsnAlaProGluTyrValGluTrpHisSerAla 268
 Db 1461 TATCCGATG-----GATATGGTGCCTTTTCCCGCAGCGC 1493
 QY 269 GluAsnLeuPheHisSerIleTyrAsnAspAspLysGlnLysLeu----- 284
 Db 1494 GATAATATTTTGCAGCGGTGTAAACCGCATGGCTATACCTGTGTGGCGCAATAACAATTTGTAC 1553
 QY 285 -----PheThrLeuLeuThrAsnHisArgCysThr 294
 Db 1554 GGGCGAAGCACTCGCCCGTGGGAAGCAGCCCTACCTGTGTGGCGCAATAACAATTTGTAC 1613
 QY 295 ArgLeuAlaLeuThrGlyGlyTyrGluLysLysIleGlyAlaAlaGluValArgAsnLeu 314
 Db 1614 CCATTTGGCCGATCGGTGCGATGTATAACGGCATCCCATGTTGAA----- 1661
 QY 315 LeuAlaThrArgAsnProSerSerGlnLeuAspSerTyrIleMet----- 329
 Db 1662 CCGCGCGAAGCAATGGCGGTGTGTGGCGGAGCGGTGTTTATAAAATGGACACC 1721
 QY 329 ----- 329
 Db 1722 GACAGCAATAACCGTATCACCGCGTTCACTGGCTGGATACCTCTGGCGCTTCGATCAA 1781
 QY 330 -----AlaLysValTyrValLeuAlaSerGlyAlaIleGlyAsnProGlnIleLeuTyr 347
 Db 1782 GCCACGGGAAAGCCTTCGCGTGGCTGCAATGTCATCGAGACGCGCGTGTGTGTG 1841
 QY 348 AsnSerGlyPheSerGlyLeuGlnValThrProArgAsnAspSerLeuIleProAsn--- 366
 Db 1842 -----ATGCGCGGAATGACGCTAATCCCAACGGTATCCCGACCGT 1883
 QY 367 -----LeuGlyArgTyrIleThrGluGlnProMetAlaPheCysGlnIleValLeu 383
 Db 1884 TCCGATATGTTGGCGGTAAACATGATGACCACTCCGCGTCCACTGTTCGTTCTGACG 1943
 QY 384 ArgGlnGluPheValAspSerValArgAspAspProTyrGlyLeuProTrpTrpLysGlu 403

1944 AAAGAA----- 1958
 404 AlaValAlaGlnHisIleAlaLysAsnProThr---AspAlaLeuProIleProPheArg 422
 1959 -----CTGGCAAGGTCCGGCGCAACAGCAGCTGTATGTGTCGGCTACCGCT 2003
 423 AspProGlu-ProGlnValThrProPheThrGluGluHisPro----- 437
 2004 GACGGCGACTTCCGACGCGACTACTCCGGAACAAGTGCCTCAACAATATTTCCTCGT 2063
 438 -TrpHisThrGlnIleHisArg-----AspAlaPheSerTyrGlyAlaValGlyPr 454
 2064 GTGGTCACGCGCAACGACACAGCAGCATGAAAAAGCGTGTGGTGGC-AAAGACCTGGATGA 2122
 454 oGluValAspSerArgValIle-----ValAspLeuArgTrpPheGlyAlaThrAspPr 472
 2123 GGAGATTGCGTACCAGCGAGTCCATAGCGTGTGATCTCTCCATCAGTCTGGGAACCGCTGCC 2182
 472 oGluAlaAsnAsnLeuValPheGlnAsnAspValGlnAspGlyTyrSerMetProGln 492
 2183 CGATCCGGAACCGCTCTGACGCTGAGCAAAACCGTAAAGATCCGCGATGGCTGGCCGTG 2242
 492 nProThrPheArgTyrArgProSerThrAlaSerAsnValArgAlaArgLysMetMetAl 512
 2243 CCGCGATATTTATTACG--ACGTGGGTGATTATGTCGTAAGGCGCGGAAGCTTCTCAC 2300
 512 aAspMetCysGluValAlaAlaSerAsnLeuGlyTyrLeuProThrSerProProGlnPh 532
 2301 GCGCAGCTGGGACATATCGGCGAGCTGT-----TTGACGCCAAA-GAAT 2344
 532 e---MetAspProGlyLeuAlaLeu-----HisLeuAlaGlyThrThrArgIleGlyPh 549
 2345 CACTATCAGTCAAGCGCTGAATGCCAATAACACATATGCGTGGGTATCATGGGAAA 2404
 549 eAspLysAlaThrThrValAlaAspAsnAsnSerLeuValTrpAspPheAlaAsnLeuTy 569
 2405 AAATGCAAAAGAGCGGTGTGGACGCAATTTGTCGGCATTTGACCATGAAAATCTCTG 2464
 569 rValAlaGlyAsnGlyThrIleArgThrGlyGlyGluAsnProThrLeuThrSerMe 589
 2465 GCTGCCCGGGGAGCGGCATCCCTCCGCCAGCGTGGTGAACAGTACGTACCTGACCATGGC 2524
 589 tCysHisAlaIleLysSerAlaArgSerIleIleAsnThrLeuLysGly 605
 2525 GCGCTGGGCTGAAGCGCGCTCATGACATCAGCTGGCGCATGAAGGG 2573

RESULT 11
 AAX57912
 ID AAX57912 standard; DNA; 5187 BP.
 XX
 AC AAX57912;
 XX
 DT 15-JUL-1999 (first entry)
 XX
 DE G. oxydans D-sorbitol dehydrogenase coding sequence.
 XX
 KW D-sorbitol dehydrogenase; L-sorbose; 2-keto-L-gulonic acid; precursor;
 KW L-ascorbic acid production; ss.
 XX
 OS Gluconobacter oxydans.
 XX
 PN WO9920763-A1.
 XX
 PD 29-APR-1999.
 XX
 PF 13-OCT-1998; 98WO-JP04612.
 XX
 PR 17-OCT-1997; 97JP-0285280.
 XX
 PA (FUJI) FUJISAWA PHARM CO LTD.
 XX
 PI Ishii Y, Noguchi Y, Saito Y, Soeda S, Yoshikawa K;

XX WPI: 1999-302741/25.
 XX Gene group for D-sorbitol dehydrogenase, useful for simple
 PT large-scale production of L-sorbose or 2-keto-L-gulononic acid as
 PT precursor for L-ascorbic acid
 XX Claim 21; Page 60-62; 83pp; Japanese.
 XX This sequence encodes the D-sorbitol dehydrogenase of the
 CC invention. Cells transformed with a vector containing DNA encoding
 CC the dehydrogenase can be used to produce L-sorbose or 2-keto-L-gulononic
 CC acid as precursor for simple large-scale L-ascorbic acid production.
 XX Sequence 5187 BP; 1251 A; 1390 C; 1297 G; 1249 T; 0 other;
 SQ

Alignment Scores:

Pred. No.:	1.22e-07	Length:	5187
Score:	206.00	Matches:	149
Percent Similarity:	30.44%	Conservative:	86
Best Local Similarity:	19.30%	Mismatches:	231
Query Match:	6.27%	Indels:	307
DB:	20	Gaps:	33

US-09-856-327-2 (1-618) x AAX57912 (1-5187)

QY 8 MetLeuArgAspTyrProArgSerMetGlnIleAsnGly----- 20
 DB 1144 ATGATGCGGTCTTCCCAAGACTTATGCGACGAGACCCTTCTACTGGACTGAAAAGC 1203
 QY 21 -----GlnIleProLysAsnAlaIleHisGluThrTyr----- 31
 DB 1204 CACCAGTCGTTGAGAGCGCAACAGCGCGCCGCGACTGTCTCCATCGGAA-TATGTCGA 1262
 QY 32 -----GlyAsnAspGlyValasp 37
 DB 1263 GAATCCCAAGTAGAACGGATGTTATTCATGAGTCTTCAATTCCTTCGTCGACAGAT 1322
 QY 38 ValPheIleAlaGlySerGlyProIleGlyAlaThrTyrAlaLysLeuCysValGluAla 57
 DB 1323 GTCGATGATCGGTGATCGGCGTCGACGGGCCAGTATGCGCAACGAACCTTCCGAGAGCC 1382
 QY 58 GlyLeuArgValMetValGluIleGlyAlaAlaAspSerPheTyrAlaValAla 77
 DB 1383 GGCCTCTCCGTCATCGTCTTGAAGCGCGCC----- 1415
 QY 78 GluGluGlyThrAlaValProTyrValProGlyTyrHisLysLysAsnGluIleGluPhe 97
 DB 1416 -----CGGATCGACCGC 1427
 QY 98 GlnLysAspIleAspArgPhe---ValAsnValIleLysGlyAlaLeuGlnGlnValSer 116
 DB 1428 CAGCATATCTTGAAATTTCCGCCACCAGGAAACAGGGGACATACCACTTCCCTAC 1487
 QY 117 ValProValArgAsnGlnAsnValProThrLeuAspProGlyAlaTrpSerAlaProPro 136
 DB 1488 CCACCCGTG-----CCTTGGCGATGATCCG 1514
 QY 137 GlySerSerAlaIleSerAsnGly-----LysAsnProHisGlnArgGluPhe 152
 DB 1515 CCTGATAGGCTCTCCCAATGGCTATCTGCATACGACCGGACCTGACGGTCTCCGAT 1574
 QY 153 GluAsnLeuSerAlaGluAlaValThrArgGlyValGlyGlyMetSerHisTrpThr 172
 DB 1575 CAG-----CAGGGTATCTGCTGTGTCGGGGGACACCTGGCATGGGCA 1622
 QY 173 CysSerThrProArgIleHisPro-----ProMetGluSerLeuProGlyIleGly 189
 DB 1623 GGATGTCCTGGCGGTATCTCCCTCTGACTTCGATTTACATTCCTCCGATATGGCGTTGGC 1682
 QY 190 ArgProLysLeuSerAsnAspProAlaGluAspAspLysGlyLutProAsnGluLeuTyrSer 209
 DB 1683 CGC-----GACTGGGCCCATCAAGTACGATGATCTCTGGAGCCATTCTACTAT 1727

QY 210 GluAlaGluArgLeuIleGlyThrSerThrLysGluPheAspGluSerIleArgHisThr 229
 DB 1728 CAGGCCGAAGTCATGATGGCGTGGCAGCCCTAACATGGAT----- 1769
 QY 230 LeuValLeuArgSerLeuGlnAspAlaTyrLysAspArgGlnArgIlePhe-----Arg 247
 DB 1770 -----GTGATGACCTGGGATCTCCACGATCTCAACAATTCACCGATGAAG 1814
 QY 248 ProLeuProLeuAlaCysHisArgLeuLysAsnAlaProGluTyrValGluTrpHisSer 267
 DB 1815 GAAGTACCCCTGTCC----- 1829
 QY 268 AlaGluAsnLeuPheHisSerIleTyrAsnAspAspLysGlnLysLysLeuPheThrLeu 287
 DB 1830 -----TATGGCGGGATCAGTTTCGCACACTGATCCATGAG 1865
 QY 288 LeuThrAsnHisArg----- 292
 DB 1866 AAGACGAATTACCGCGTGTTCACGAGCCACAGCCGCTAACACTCGCCCTTATGACAAG 1925
 QY 293 -----CysThrArgLeuAlaLeuThr 299
 DB 1926 CGCCCACTGTGAGGGCAACAACACTGATGCTGTCGATCGGCGGCGATGTAC 1985
 QY 300 GlyGly-----TyrGluLysLysIleGlyAlaAlaGluValArgAsnLeu 314
 DB 1986 AACGGAATTCACCTCGTCAATCATCGGAGCAGCAGCGCCGCTATATTCGATGCG 2045
 QY 315 LeuAlaThrArgAsnProSerSerGlnLeuAspSerTyrIleMet----- 329
 DB 2046 GTTGTCTACCGACTGGAGACCAGCCAGCAACAAGAGTCTGTCGCCCTAAATATTAT 2105
 QY 330 -----AlaLysValTyrValLeuAlaSerGlyAla 339
 DB 2106 GATCCCGATAAGAAATTCATGCTGTCACCGGTAAGTCTTCGTGCTGCTCGCACTGC 2165
 QY 340 Ile-----GlyAsnProGlnIleLeuTyr 347
 DB 2166 ATTGAGAGTGCCAAAGCTGCTCTGCTGCTGCGCGATGATGATGATGATGATGATGAT 2225
 QY 348 AsnSerGly-----PheSerGlyLeuGlnValThr----- 357
 DB 2226 AACATTCAGATCAGGTTGGTGGGAACATGATGATGATGATGATGATGATGATGATGAT 2285
 QY 358 ---ProArgAsnAspSerLeuIleProAsnLeuGlyArgTyrIleThrGluGlnProMet 376
 DB 2286 ATGACCGGAACAGCTCTCTGTGGCGGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2333
 QY 377 AlaPheCysGlnIleValLeuArgGlnGluPheValAspSerValArgAspProTyr 396
 DB 2334 -----AGCATTATCGACTCGTTCGTGACGGC----- 2360
 QY 397 GlyLeuProTrpTrpLysGlu-----AlaValAlaGlnHisIleAlaLysAsnProThr 414
 DB 2361 -----CCATGGCGGAGGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2405
 QY 415 AspAlaLeuProIleProPheArgAspProGluProGlnValThrThrProPheThrGlu 434
 DB 2406 -----GATAATCAGGTCGACTTCGCAACGGGTCTG----- 2435
 QY 435 GluHisProTrpHisThrGlnIleHisArgAspAlaPheSerTyrGlyAlaValGlyPro 454
 DB 2436 -----GCGATTGCCAGGGCTATTCGCGAAA 2462
 QY 455 GluValAspSerArgValIleValAsp-----LeuArgTrpPhe----- 467
 DB 2463 GAGCTGGAAGACGACATCGGTTATGGCTCTCTCATGCGGTTCGCTCTCTTCAGCCATAAC 2522
 QY 468 ---GlyAlaThrAspProGluAlaAsnAsnLeuValPheGlnAsnAspValGlnAsp 486
 DB 2523 GAAGGCTGCGGACCCGAC-----AACCGGCTGACACTGAGCAACACATAAAGAC 2576

Qy	487	GlyTyrSerMetProGlnProThrPheArgTyrArg-----ProSerThrAlaSer	503
Db	2577	GTTCGTGGCATCTCTCAACCGGAAGTCTATTACAAGTTCCTCCGAGTACACAGTGAAGAGT	2636
Qy	504	AsnValArgAlaAtrGlySerMetMetAlaAspMetCysGluValAlaSerAsnLeuGly---	522
Db	2637	TGTGACCATACCAAGGAGCTGTTCAAGGAAGTGTGATGCTCTGATGAGTGGTACTGATCCT	2696
Qy	523	-----GlyTyrLeuProThrSerProProGlnPheMetAspProGlyLeuAla	538
Db	2697	CAATGGACAAGGGTTACTTCCCG-----CAG	2723
Qy	539	LeuHisLeuAlaGlyThrThrArgIleGlyPheAspLysAlaThrThrValAlaAspAsn	558
Db	2724	TGCCATCCGTCCGGCAGCAGATCATGGGAACAGACCCCAATTCGGTCGTGACGGT	2783
Qy	559	AsnSerLeuValTrpAspPheAlaAsnLeuTyrValAla-----	571
Db	2784	GAGTGC CGCACCCATGACACCAAGAAACCTGTTGTGCCAGATCAGCGGTCTTCTCTTCG	2843
Qy	572	---GlyAsnGlyThrIleArgThrGlyPheGly-----	581
Db	2844	GTCGGTACAGCAATATACCTGACCATGGCGGCTTCGCGGTTCGCGTTGCAGCATCC	2903
Qy	582	-----GlyTyrLeuProThrPheArgTyrArg-----GluAsnProThrLeuThr	587
Db	2904	CTGAAAAAGGAGATGCTTTCATGCTGAGGGGAATAAAGCCGGAATACGCCGCTCTTCT	2963
Qy	588	SerMetCysHisAla-----IleLysSerAlaArgSerIleIleAsnThrLeu	603
Db	2964	GCCAGCTGCCATAGCTTCGGGTGTCTCTGTTTCGGCGCGCAGTCAGC-----	3008
Qy	604	LysGlyGlyThrAspGlyLysAsnThrGlyGluHis	615
Db	3009	-----GAGGCAGAGGATCAGGCCACCAC	3032
RESULT 12			
ID	AAX57909		
XX	AAX57909 standard; DNA; 1632 BP.		
AC	AAX57909;		
XX	15-JUL-1999 (first entry)		
XX	G. oxydans D-sorbitol dehydrogenase coding sequence.		
DE	D-sorbitol dehydrogenase; L-sorbose; 2-keto-L-gulononic acid; precursor;		
KW	L-ascorbic acid production; ss.		
XX	Gluconobacter oxydans.		
OS	WO9920763-A1.		
PN	29-APR-1999.		
PD	13-OCT-1998; 98WO-JP04612.		
XX	17-OCT-1997; 97JP-0285280.		
PF	(FUJI) FUJISAWA PHARM CO LTD.		
XX	Ishii Y, Noguchi Y, Saito Y, Soeda S, Yoshikawa K;		
PI	WPI: 1999-302741/25.		
DR	P-PSDB: AAY14049.		
XX	Gene group for D-sorbitol dehydrogenase, useful for simple		
PT	large-scale production of L-sorbose or 2-keto-L-gulononic acid as		
PT	precursor for L-ascorbic acid		
XX	Claim 14; Page 52-54; 83pp; Japanese.		
PS	This sequence encodes the D-sorbitol dehydrogenase of the		
XX			

```
Db 628 AAGCGCCCAACCTGTGTAGGCGCAACAACAACTGCATGCGCATCTGTCCGATCGGGCGGATG 687
Qy 299 ThrGlyGly-----TyrGluLysLysIleGlyAlaAlaGluValArgAsn 313
Db 688 TACAACGGAATTAAGTCAATCAATCGGAGCAGCAGCGCCGCTATTATTCGGAAT 747
Qy 314 LeuLeuAlaThrArgAsnProSerSerGlnLeuAspSerTyrIleMet----- 329
Db 748 GCGGTGTCTACCGACTGGACCGCAGCCGACAGCAAGAGGTCTGCCCGTAAATAT 807
Qy 330 -----AlaLysValTyrValLeuAlaSerGly 338
Db 808 TAGCATCCCGATAAGAATTTCTATCGTGTACCGGTAAAGTTCTCGTGTGCTCGCGCAC 867
Qy 339 AlaIle-----GlyAsnProGlnIleLeu 346
Db 868 TGCATTGAGAGTCCCAAGCTCCTCTGTGTCCCGCGATGACAAAATCCCGGGGCAAT 927
Qy 347 TyrAsnSerGly-----PheSerGlyLeuGlnValThr 357
Db 928 GCCAACAGCTTCAGATCAGGTTCGGAACATGATGGATCACACGGGCGGTACAGCTCTCG 987
Qy 358 -----ProArgAsnAspSerIleProAsnLeuGlyArgTyrIleThrGluGlnPro 375
Db 988 TTTATGACGGAAACGACTCTCTGTGGCCGGGTCTGTCTCTGCTGAC----- 1038
Qy 376 MetAlaPheCysGlnIleValLeuArgGlnGluPheValAspSerValArgAspAspPro 395
Db 1039 -----AGCATATTCGACTCGTTTCGTTCGTGACGGC--- 1065
Qy 396 TyrGlyLeuProTyrTrpLysGlu-----AlaValAlaGlnHisIleAlaLysAsnPro 413
Db 1066 -----CCATGGCGGAGGAGCGTGTGCGTATCTTGTGCATATGTTGAC----- 1110
Qy 414 ThrAspAlaLeuProIleProPheArgAspProGluProGlnValThrThrProPheThr 433
Db 1111 -----GATAATCAGGTCTGACTTCGCAACGCGTCTG--- 1140
Qy 434 GluGluHisProTrpHisThrGlnIleHisArgAspAlaPheSerTyrGlyAlaValGly 453
Db 1141 -----GCGATTGCCAAGGGCTATGTCGGG 1164
Qy 454 ProGluValAspSerArgValIleValAsp-----LeuArgTrpPhe----- 467
Db 1165 AAAGAGCTGGAAGACAGACATCGTTATGGCTCTCTCATGCGTTCGTTCTTCAGCCAT 1224
Qy 468 -----GlyAlaThrAspProGluAlaAsnAsnLeuLeuValPheGlnAsnAspValGln 485
Db 1225 AACGAAGGCATTGCGGACCCGAC-----AACCGCTGACACTGAGCAAAACACATAAA 1278
Qy 486 AspGlyTyrSerMetProGlnProThrPheArgTyrArg-----ProSerThrAla 502
Db 1279 GACGTTCTGGGCATTCCTCACCCGAGCTATTACAGAGCTCCGAGTACACAGTGAAG 1338
Qy 503 SerAsnValArgAlaArgLysMetMetAlaAspMetCysGluValAlaSerAsnLeuGly 522
Db 1339 AGTTGTGACCATACCAAGGAGCTGTTCAAGAACTGATGGCTCTCATGAGTGGTACTGAT 1398
Qy 523 -----GlyTyrLeuProThrSerProGlnPheMetAspProGlyLeu 537
Db 1399 CCTCAATGGCAAAAGGTTACTTCCG----- 1425
Qy 538 AlaLeuHisLeuAlaGlyThrThrArgIleGlyPheAspLysAlaThrThrValAlaAsp 557
Db 1426 CAGTGCCCATCGTGGCGGACGATCATGGAGACAGACCCACCAATTCGGTCTGTCAC 1485
Qy 558 AsnAsnSerLeuValTrpAspPheAlaAsnLeuTyrValAlaGlyAsnGlyThrIleArg 577
Db 1486 GGTGAGTCCCGCACCCATGACACAGAACTGTTGTGTGGCAGATCAGCGGTCTTCTCT 1545
Qy 578 ThrGlyPheGlyGluAsnProThrLeuThrSerMetCysHisAlaIleLysSerAlaArg 597
Db 597 ----- 1605
```

```
Db 1546 TCGGTGCGTACAGGCAATATCACCTGACCATTTGGCGGCTTCGGCTTCGAGCA 1605
Qy 598 SerIle 599
Db 1606 TCCCTG 1611
RESULT 13
AAAX25345
ID AAX25345 standard; DNA; 4665 BP.
AC AAX25345;
XX 19-JUL-1999 (first entry)
XX Membrane-bound gluconate dehydrogenase gene.
DE Gluconate dehydrogenase; membrane-bound; 2-keto-gluconate;
KW vitamin C; ascorbic acid; cytochrome c; ss.
XX Erwinia cyripedii.
OS Key Location/Qualifiers
PH CDS 258..290
FT /*tag= a
FT /product= "subunit III"
FT /note= "specifically claimed in Claim 10, encodes
FT AAY05721"
FT sig_peptide 258..380
FT /*tag= b
FT mat_peptide 381..917
FT /*tag= c
FT RBS 925..929
FT /*tag= d
FT /*standard_name= "Shine-Dalgarno"
FT CDS 934..2781
FT /*tag= e
FT /product= "subunit I (dehydrogenase)"
FT /note= "specifically claimed in Claim 12; encodes
FT AAY05719"
FT CDS 933..998
FT /*tag= f
FT mat_peptide 999..2778
FT /*tag= g
FT RBS 2781..2784
FT /*tag= h
FT /*standard_name= "Shine-Dalgarno"
FT CDS 2793..4118
FT /*tag= i
FT /product= "subunit III (cytochrome c)"
FT /note= "specifically claimed in Claim 14; encodes
FT AAY05720"
FT sig_peptide 2793..2849
FT /*tag= j
FT mat_peptide 2850..4115
FT /*tag= k
FT WO9915673-A1.
PN 01-APR-1999.
XX 25-SEP-1998; 98WO-KR00296.
XX 25-SEP-1997; 97KR-0048802.
XX (KOAD ) KOREA ADV INST SCI & TECHNOLOGY.
PA (KOMI-) KOREA MICROBIAL TECHNOLOGY INC.
XX Pan JG, Yum DY;
XX WPI; 1999-254718/21.
DR P-FSDB; AAY05719, AAY05720, AAY05721.
XX New isolated gluconate dehydrogenase
```



```

Db      2520 GGGTACCACCTTCGATACCACCGGTGTATCAA-----ACCCACGATAT 2561
Qy      541 uAlaGlyThrThrArgIleGlyPheAspLysAlaThrThrValAlaAspAsnSerLe 561
Db      2562 GAGCGCGGGGGATCATGGGTGAAGATCCGAAACCCAGCGAGTGAACCGTATTTCGA 2621
Qy      561 uValTrpAspPheAlaAsnLeuTyrValAlaGlyAsnGlyThrIleArgThrGlyPheG1 581
Db      2622 GAGCTGGGATGCGCGAAGCGTGTGTGCGGGTGCCTCCGGTCCCGCAGGGTCTGGG 2681
Qy      581 yGluAsnProThrLeuThrSerMetCysHisAlaIleLysSerAlaArgSerIle-IleA 601
Db      2682 CTACAACCCGACCGCATCGGTGGCGGCACCTGACCTACTGCTGCGGAAAGCCATCCGTGA 2741
Qy      601 snThrLeu-----LysGlyGlyThrAspGlyL 610
Db      2742 ACAGTATCTGAAGAACCCAGGTCCACTGGTGCAGGCATGAAGAAACGGCGATGATGAAA 2801
Qy      610 ysAsnThrGly 613
Db      2802 AGCATTCGTGGC 2812

RESULT 14
ABL52744
ID      ABL52744 standard; DNA; 1689 bp.
XX
AC      ABL52744;
XX
DT      01-JUL-2002 (first entry)
XX
DE      Glucose-3-dehydrogenase encoding sequence.
XX
KW      Glucose-3-dehydrogenase; G3DH; enzyme; ds.
XX
OS      Halomonas sp. alpha-15.
XX
FH      Key
FT      CDS
FT      Location/Qualifiers
FT      /*tag= a
FT      /product= "glucose-3-dehydrogenase"
XX
PN      JP2002017372-A.
XX
PD      22-JAN-2002.
XX
PF      30-JUN-2000; 2000JP-0237709.
XX
PR      30-JUN-2000; 2000JP-0237709.
XX
PA      (HAYA/) HAYADE K.
XX
DR      WPI; 2002-221711/28.
DR      P-PSDB; ABB09452.
XX
PT      Glucose-3-dehydrogenase and its preparation -
PS      Claim 5; Page 9-10; 16pp; Japanese.
XX
CC      The invention relates to Glucose-3-dehydrogenase (G3DH) enzyme. The
CC      methods of the invention are used for the preparation of G3DH. The
CC      current sequence represents the glucose-3-dehydrogenase encoding
CC      sequence.
XX
SQ      Sequence 1689 BP; 380 A; 446 C; 508 G; 355 T; 0 other;

Alignment Scores:
Pred. No.:      1.69e-05      Length:      1689
Score:          174.00      Matches:      141
Percent Similarity: 33.18%      Conservative: 79
Best Local Similarity: 21.27%      Mismatches: 242
Query Match:      5.30%      Indels:      201
DB:              24      Gaps:      33

```

```

US-09-856-327-2 (1-618) x ABL52744 (1-1689)
Qy      37 AspValPheIleAlaGlySerClyProIleGlyAlaThrTyrAlaLysLeuCysValGlu 56
Db      19 GATGCCATCGGTGGTGGCTCAGGCATTTAGCGCGGTGGTGGCTGCTAAAGAGTAACCCGAG 78
Qy      57 AlaGlyLeuArgValValMetValGluIleGlyAlaAlaAspSerPheTyrAlaValAsn 76
Db      79 AAGGGCTAAAGGTTTTCCTGCTT----- 102
Qy      77 AlaGluGluGlyThrAlaValProTyrValProGlyTyrHisLys----- 91
Db      103 ---GAGCGTGGCGCGCAATATTAGCAGCATTAAGAGGATTACCAATGCGGACAAAGAGGCG 159
Qy      92 -----LysAsnGluIleGluPheGlnLysAspIleAspArgPheValAsn 106
Db      150 TGGGATTACCCCGCCACCGTAATGAG---CCAACCCAGGAATGATCGCTAATATAT----- 210
Qy      107 ValIleLysGlyAlaLeuGlnValSerValProValArgAsnGlnAsnValPro--- 125
Db      211 -----CCTGTGCTCAAGCGCGACTACCCCTT 237
Qy      126 -----ThrLeuAspProGlyAlaTrpSerAlaProProGlySerSerAlaIleSer 142
Db      238 AATGAAGCCACCCCTG---GGCATGTGGCG-----GAT 267
Qy      143 AsnGlyLysAsnProHis-----GlnArgGluPheGluAsnLeuSerAlaGluAlaVal 160
Db      268 GAACAGGCCAATCCCTACGTAGAGAGAAAGCGCTTCGACTGGTTT----- 312
Qy      161 ThrArgGly-----ValGlyClyMetSerThrHisTrpThrCysSerThrProArgIle 178
Db      313 ---CGCGCTACACGAGTAGCGGCGCTCTCTGCTTGGGGTGGCGCAAGCTACCGGCTC 369
Qy      179 HisProMetGluSerLeuProGlyIleGlyArgProLysLeuSerAsnAsp----- 196
Db      370 AGCCCG-----ATGGATTTTGAGGCCAACAGCGGAGGAGCATCGCCATGATTGGCGC 423
Qy      197 ---ProAlaGluAspAspLysGluTrpAsnGluLeuTyrSerGluAlaGluArgLeuIle 215
Db      424 ATTCGCTACGAAGATCTCGCCCGCTGG-----TATGACTATGTGGAGCGCTTGGC 474
Qy      216 Gly---ThrSerThrLysGluPheAspGluSerIleArgHisThrLeuValLeuArgSer 234
Db      475 GGTATCGCAGCGACCCAGGAG-----GGCTGGATATC 507
Qy      235 LeuGlnAspAlaTyrLysAspArgGlnArgIlePheArgProLeuProLeuAlaCys--- 253
Db      508 CTCCCGATGGC-----GAGTTTCTCGCGCAATACCGCTTAACCTGTGTG 552
Qy      254 -----HisArgLeuLysAsnAlaProGluTyrValGluTrpHisSerAla 268
Db      553 GAAGAGGATCGGCCAACGCGTATTAAAGAGGCG-----TTTGGTGGCGAG 597
Qy      269 GluAsnLeuPheHisSer-----IleTyrAsnAspAspLysGlnLysLysLeu 284
Db      598 CGTCACTTATCCACAGCGGGTGGCCCAATATTACCCAGCCAAAGCCAGACCAACCGC 657
Qy      285 PheThrLeuLeuThrAsnHisArgCysThrArgLeuAlaLeuThrGlyGlyTyrGluLys 304
Db      658 GTCAATTGCCAATACCGAAATAAATGCTGGCTGGGTGCTCCCTTATGTGGTGGCTAT----- 711
Qy      305 LysIleGlyAlaAlaGluValArgAsnLeuLeuAlaThrArgAsn-----ProSer 321
Db      712 -----TTCAGACCCCGCTCGCTACGCTTCCCGCG 741
Qy      322 SerGlnLeuAspSerTyrIleMetAlaLysValTyrValLeuAlaSerGlyAlaIleGly 341
Db      742 CGGTGGCCACAGGTAATCTACGTTACGGCATCTCGATTGTCAGC----- 789
Qy      342 AsnProGlnIleLeuTyrAsnSerGly-PheSerGlyLeuGlnValThrProArgAsnAs 361

```


[illegible]

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 24, 2003, 01:52:53 ; Search time 4438 Seconds

(without alignments)

11533.399 Million cell updates/sec

Title: US-09-856-327-1

Perfect score: 2106

Sequence: 1 atcagccatgtctctctca.....aaaaaaaaaaaaaaaaaaaaa 2106

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.*

1: em_estba.*

2: em_esthum.*

3: em_estin.*

4: em_estmu.*

5: em_estov.*

6: em_estpl.*

7: em_estro.*

8: em_htc.*

9: gb_est1.*

10: gb_est2.*

11: gb_htc.*

12: gb_est3.*

13: gb_est4.*

14: gb_est5.*

15: em_estfun.*

16: em_estom.*

17: em_gss_hum.*

18: em_gss_inv.*

19: em_gss_pln.*

20: em_gss_vrt.*

21: em_gss_fun.*

22: em_gss_nam.*

23: em_gss_mus.*

24: em_gss_pro.*

25: em_gss_rod.*

26: em_gss_phg.*

27: em_gss_vrl.*

28: gb_gsl.*

29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	97.4	4.6	630	14	CA349327
2	96.2	4.6	1201	13	BX446296
3	95.8	4.5	553	14	CB054499
4	95.4	4.5	336	12	BM884782

C	5	95.4	4.5	480	9	AI392214	AI392214	NCSP1F9T7
C	6	95.2	4.5	871	9	AL514867	AL514867	AL514867
	7	95	4.5	271	13	BU531092	AGENCOURT	BU531092
	8	95	4.5	503	11	AY109012	2ea mays	AY109012
	9	94.8	4.5	509	13	BQ276784	AGENCOURT	BQ276784
	10	94.2	4.5	165	12	BI746603	rm29d01.y	BI746603
	11	94.2	4.5	286	10	BG736597	rk59f08.y	BG736597
	12	94	4.5	974	13	BX327262	BK327262	BX327262
	13	93.8	4.5	318	12	BI941623	sc80a12.y	BI941623
C	14	93.8	4.5	381	13	BX425197	BX425197	BX425197
C	15	93.8	4.5	417	13	BQ526732	NISC_no18	BQ526732
	16	93.6	4.4	263	12	BM900884	rc42f02.y	BM900884
	17	93.6	4.4	809	14	-CA453483	AGENCOURT	CA453483
	18	93.6	4.4	911	13	BX380015	BX380015	BX380015
C	19	93.6	4.4	1104	9	AL513809	AL513809	AL513809
	20	93.4	4.4	162	9	AW395255	SH45911.y	AW395255
C	21	93.4	4.4	1201	9	AL513901	AL513901	AL513901
	22	93.2	4.4	436	14	CA799440	sat33g04	CA799440
	23	93.2	4.4	484	12	BI680498	458776 MA	BI680498
	24	93.2	4.4	949	13	BQ433605	AGENCOURT	BQ433605
	25	93.2	4.4	1017	14	CB946057	AGENCOURT	CB946057
	26	93	4.4	248	13	BU589063	AGENCOURT	BU589063
	27	93	4.4	600	12	BG926622	HNC56-1-H	BG926622
	28	93	4.4	685	9	AV714975	AV714975	AV714975
	29	93	4.4	936	13	BU850422	AGENCOURT	BU850422
	30	93	4.4	1334	12	BM555708	AGENCOURT	BM555708
	31	92.8	4.4	299	14	CA688814	wlm96.pk0	CA688814
	32	92.8	4.4	344	12	BM514150	xx96e09.y	BM514150
	33	92.8	4.4	593	12	BM514170	xx97a03.y	BM514170
	34	92.6	4.4	262	10	BE722980	192270 MA	BE722980
	35	92.6	4.4	314	14	CA802494	sau37b11	CA802494
	36	92.6	4.4	337	12	BI142697	kt44f06.y	BI142697
C	37	92.6	4.4	409	13	BX425137	BX425137	BX425137
	38	92.6	4.4	421	10	BG362036	9b50f09.y	BG362036
	39	92.6	4.4	586	14	CA370380	650538 NC	CA370380
	40	92.6	4.4	778	13	BU533506	AGENCOURT	BU533506
	41	92.4	4.4	347	13	BU834850	T066E04 P	BU834850
	42	92.4	4.4	434	14	CA849260	k109h12.y	CA849260
	43	92.4	4.4	620	14	CA353533	625052 NC	CA353533
	44	92.4	4.4	844	13	BU908877	AGENCOURT	BU908877
	45	92.2	4.4	308	10	BE633260	uv74g09.y	BE633260

ALIGNMENTS

RESULT 1

CA349327

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CA349327 619706 NCCGWA lrt Oncorhynchus mykiss cdna clone lrt140M18_C_G09
5', mRNA sequence.

CA349327 1 GI:24594498

EST.

Oncorhynchus mykiss (rainbow trout)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Actinopterygii; Neopterygii; Teleostei; Euteleostei;

Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.

1 (bases 1 to 630)

Rexroad,C.B. and Keele,J.W.

Sequence analysis of a rainbow trout normalized cdna library

Unpublished

Contact: Rexroad CE

USDA, ARS, National Center for Cool and Cold Water Aquaculture

11876 Leetown Road, Kearneysville, WV 25430, USA

Tel: 304 724 8340 x2129

Fax: 304 725 0351

Email: crexroad@nccwa.ars.usda.gov

Single pass sequencing. Bases called with phred v0.020425.c and

trimmed with the aid of the trim_alt option. Vector identified by

cross_match v0.990329.

Seq primer: AGCGGATAACAATTTTCACACAGGA.

PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150560.
PR 26-AUG-1999; 99US-0150584.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.

PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Alignment Scores:
Pred. No.: 0.000548 Length: 2484
Score: 159.50 Matches: 130
Percent Similarity: 33.58% Conservative: 97
Best Local Similarity: 19.23% Mismatches: 224
Query Match: 4.86% Indels: 225
DB: 21 Gaps: 30

US-09-856-327-2 (1-618) x AAC36995 (1-2484)

OY 16 MetcInlleasnnglyInlleProlysaSnAlaIleHsgIurHrYrGlyAsnAspGly 35
DB 728 TTACCATCTGCACACATCTCTACGAAAGGTTCAT---GTTCGTAGACGACGAT 784
OY 36 ValaspYalphe-----IlelaaglySerGlyProIleGlyAla 48
DB 785 GACAACTGTAACAGATCAGATCGCGTTGTCTAGCTCAGCAAGTGCAGAGCT 844
OY 49 ThrTyralalysleuCyValaGluAlaGlyleuArgValaIleMetValaGluIleGlyAla 68
DB 845 GTTCACGCTGCACAACTGATTAAGCTGATGAAGCTTGTCTAGAGAAAGGA--- 901
OY 69 AlaAspserPheTyralaValaIlelaaglyGluGlyThrAlaValProTyValProGly 88

```

Db      902 -----AACTACTTTACAGCTCATGCTATTCGGGCTAGAGAGTTCCGCTAGCTCGAG 955
QY      89 TyrHisLysLysAsnGluIleGluPheGlnLysAspIleAspArgPheValAsnValIle 108
Db      956 TTATACGAAAAAGC----- 970
QY      109 LysGlyAlaLeuGlnGlnValSerValProValArgAsnGlnAsnValProThrLeuAsp 128
Db      971 ---GGCCTTTGACTACGGTT----- 988
QY      129 ProGlyAlaTrpSerAlaProProGlySerSerAlaIleSerAsnGlyLysAsnProHis 148
Db      989 -----GATGGGAAA----- 997
QY      149 GlnArgGluPheGluAsnLeuSerAlaGluAlaValThrArgGlyValGlyMetSer 168
Db      998 -----TTTATGCTCTGGCTGGTCACTGTGGA-----GGAGTACAGCT 1039
QY      169 ThrHisTrpThrCysSerThrProArgIleHisProPheMetGluSerLeuProGlyIle 188
Db      1040 GTTAATGCTGTCATCT-----ATAAGACACCGCATCATGTTTG----- 1081
QY      189 GlyArgProLysLeuSerAsnAspProAlaGluAspAspLysGluTrpAsnGluLeuTyr 208
Db      1082 -----CAGCAATG----- 1090
QY      209 SerGluAlaGluArgLeuIleGlyThrSerThrLysGluPhe-----AspGlu 224
Db      1091 TCTGAAGGAGTAAAGATCAAGTTTGTGTACCAAGAAATATCAGCTCATGATGATGA 1150
QY      225 ---SerIleArgHisThrLeuValLeuArgSerLeuGlnAspAlaTyrLysAspArgGln 243
Db      1151 GTTAGATGAAGATTGCTGTACGGAAGAGATGTCAAAACACGGGTTTCAGAT----- 1204
QY      244 ArgIlePheArgProLeuProLeuAlaCysHisArgLeu----- 256
Db      1205 -----CAGGTTCTCGGAAGGGTCCGAGAGCTCGTTTGACAGGTAGAAATCAGTT 1255
QY      257 ---LysAsnAlaProGlu-----TyrValGluTrpHisSer 267
Db      1256 CCAAGGAACCTACACAGAGATCATCTGTGCTGTGCGGATATGAGATGAGAGAGAGA 1315
QY      268 AlaGluAsnLeuPheHisSerIleTyrAsnAspAspLysGlnLysLysLeuPheThrLeu 287
Db      1316 GCCAAGAACGGGACAGATCAAACTGGTTGTTGATGCTGTAGAGATGTCAGATGATC 1375
QY      288 LeuThrAsnHisArgCysThrArgLeuAlaLeu-----ThrGlyGlyTyrGlu 303
Db      1376 CTGACAGGGATCAAAAGCTGAGAGATTTGTATTAGTAGACACACTAGTAGCAATGAG 1435
QY      304 LysLys-----IleGlyAlaAlaGluValArgAsnLeuLeuAlaThrArg 318
Db      1436 AGAAAGAAACGATGTGTGGAGATGTTGCTACTGCTGTGAGGAAAGATGGGAAGAG 1495
QY      319 AsnProSerSerGlnLeuAspSerTyrIleMetAlaLysValTyrValLeuAlaSerGly 338
Db      1496 -----TTTATTATCGAAGCTCGAGTAGTAACGTTTCATCCCTGGA 1534
QY      339 AlaIleGlyAsnProGlnIleLeuTyrAsnSerGlyPheSerGlyLeuGlnValThrPro 358
Db      1535 TCGCTGTAAACACCGCTGTGATGCTTTCGAGTGGGCTAAAGAAC----- 1579
QY      359 ArgAsnAspSerLeuIleProAsnLeuGlyArg----- 369
Db      1580 -----CCGAACATGGAAGAACTTAAAGCTGCACCCGGTCTCGATG 1621
QY      370 -----TyrIleThrGlnGlnProMetAlaPheCysGlnIleValLeuArgGlnGlu 386
Db      1622 ACATGGGATACTCTCCGAGAAAGATTCGATTCCTCCGGAATAATGTATGAGGAGGG 1681
QY      387 PheValAspSerValArg-----AspAspProTyrGlyLeuProTyrTrpLysGluAla 404
      :::  |||||  ::::

```

```

Db      1682 ATCATCAGCTCCGCTCCATCATATGAATGAT-----ACTGAATCT 1720
QY      405 ValAlaGlnHisIleAlaLysAsnProThrAspAlaLeuProIleProPheArgAspPro 424
Db      1721 GGATCGAAAGCAATATCTGAGAGATCCACGATAGAGA---CGAGCTTTTACGACAGGTTG 1777
QY      425 GluProGlnValThrThrProPheThrGluGlnHisProTyrPheHisThrGlnIleHisArg 444
Db      1778 AGCCCGTGGGTTTCA----- 1792
QY      445 AspAlaPheSerTyrGlyAlaValGlyProGluValAspSerArgValIleValAspLeu 464
Db      1793 -----GGACCGACCTTAAAGAGAGATGATTAAG----- 1822
QY      465 ArgTrpPheGlyAlaThrAspProGluAlaAsnAsnLeuValPheGlnAsnAspVal 484
Db      1823 -----TACGAAAGAAC-----GCTCATCTTTTGTGCTGTGGTGGGACTTA 1864
QY      485 GlnAspGlyTyrSerMet-----ProGlnProThrPheArgTyrArgProSerThrAla 502
Db      1865 GGGTCAGGTGAGGTTATGATGGAACCAAGTTACATACAGGACAACCAAGAAAGATAGA 1924
QY      503 SerAsnValArgAlaArgLysMetMetAlaAspMetCysGluValAlaSer----- 519
Db      1925 GAGAAATCTGAGGGCAGCTCTTAGACAAGCTTTCGGGGTTTCGGTTGCAGCTGTCAGATT 1984
QY      520 AsnLeuGlyGlyTyr-----LeuProThr 527
Db      1985 GAAGTCGGAACATATAGAGAGTATGACAGAGATGAAGATGAGGCAATTAATAAGA 2044
QY      528 SerProProGlnPheMetAspProGlyLeuAlaLeu----- 539
Db      2045 GCCATGGAAGAGTTCTGTACACAGAGTTGATGCTGTGGTGGGTTGGCCAAAGAGAGAG 2104
QY      540 -----HisLeuAlaGlyThrThrArgIleGlyPheAspLys 551
Db      2105 TATTGGACGACCTACTTCGGCTCACCGATGGGTAGTTGATGAGATGGAGATCACCCGG 2164
QY      552 AlaThrThrValAlaAspAsnAsnSerLeuValTrpAspPheAlaAsnLeuTyrValAla 571
Db      2165 GAGAAAGTGCCTGTGATGAGATGAGATGAGACCTGGGAAGCAGAGGGGTTGTTGTTGT 2224
QY      572 GlnAsnGlyThrIleArgThrGlyPheGlyGluAsnProThrLeuThrSerMetCysHis 591
Db      2225 GATGGAAGTATATTCATACAGCTGTGGGTTTAAACCAATGATCACCATTCACTACT 2284
QY      592 AlaIleLysSerAlaArgSerIleIleAsnThrLeuLysGlyGlyThr 607
Db      2285 GCTTACTGATCTCCTCAAAAGATAGTTGACTGCTGCAAAACAAACA 2332

```

Search completed: September 24, 2003, 06:06:14
Job time : 558 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: September 24, 2003, 05:32:15 ; Search time 6374 Seconds
(without alignments)
3966.450 Million cell updates/sec

Title: US-09-856-327-2
Perfect score: 3284
Sequence: 1 MSLSFGMRDYPKSMQING.....IINTLKGTGKNTGEHRL 618

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 20454813386 residues
Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODE=frame.p2n.model -DEV=xlh
-Q/cg22_1/USFTO.spool/US09856327/runat_23092003_153419_15506/app_query.fasta.1.775
-DB=GenBml -QFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blomsum62 -TRANS=human40.cdi -LIST=45
-DOCLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pic -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USFR=US09856327 @CG22_1.1.3963 @runat_23092003_153419_15506 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WAMP_TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenBml: *
1: gb_ba: *
2: gb_htg: *
3: gb_in: *
4: gb_om: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sts: *
12: gb_sy: *
13: gb_un: *
14: gb_vi: *
15: em_ba: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_mu: *
20: em_om: *
21: em_or: *
22: em_ov: *
23: em_pat: *
24: em_pl: *
25: em_ph: *
26: em_ro: *
27: em_sts: *
28: em_un: *

29: em_vl: *
30: em_htg_hum: *
31: em_htg_inv: *
32: em_htg_other: *
33: em_htg_mus: *
34: em_htg_pln: *
35: em_htg_rtd: *
36: em_htg_mam: *
37: em_htg_vrt: *
38: em_sy: *
39: em_htgo_hum: *
40: em_htgo_mus: *
41: em_htgo_other: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3284	100.0	2106	6 BD013176	BD013176 A novel p
2	1239.5	37.7	2064	8 AF535193	AF535193 Penitophor
3	1239.5	37.7	2069	8 AY291124	AY291124 Trameetes
4	1232.5	37.5	1869	6 E11766	E11766 cDNA of pyr
5	1232.5	37.5	1869	6 I82410	I82410 Sequence 1
6	1232.5	37.5	2046	8 D73369	D73369 Cortiolus ve
7	1232	37.5	1902	6 E32737	E32737 Panus cornu
8	1210	36.8	1869	6 AR141573	AR141573 Sequence
9	1091	33.2	1701	6 AR170485	AR170485 Sequence
10	1044	31.8	1946	6 E33665	E33665 Novel prote
11	1044	31.8	1946	6 E37379	E37379 Reagent for
12	1044	31.8	1946	6 AB043883	AB043883 Trichobac
13	243	7.4	165137	1 ANI507836	ANI507836 Arthroba
14	227	6.3	305520	1 AE016780	AE016780 Pseudomon
15	208.5	6.3	2465	1 AF430844	AF430844 Burkholde
16	208.5	6.3	2467	6 BD168939	BD168939 Novel glu
17	207.5	6.3	299700	1 AP005944	AP005944 Bradythiz
18	207	6.3	4070	1 AF068066	AF068066 Erythrina h
19	206	6.3	39821	1 AB039821	AB039821 Glucosoba
20	205.5	6.3	198050	1 AL646061	AL646061 Ralstonia
21	200.5	6.1	308601	1 CJ1168X2	CJ1168X2 Campyloba
22	198	6.0	4665	1 EC97665	EC97665 Erythrina cyp
23	198	6.0	4665	6 AR182220	AR182220 Sequence
24	191.5	5.8	1845	6 AR182218	AR182218 Sequence
25	190.5	5.8	13036	1 AE005839	AE005839 Caulobact
26	190	5.8	349050	1 AP003586	AP003586 Nostoc sp
27	186	5.7	13508	1 AE005037	AE005037 Halobacte
28	183.5	5.6	11558	1 AE004652	AE004652 Pseudomon
29	180	5.5	4349	1 AF131202	AF131202 Pantoea c
30	177	5.4	299800	1 AP005028	AP005028 Streptomy
31	176.5	5.4	213732	1 AE001862	AE001862 Deinosococ
32	174	5.3	1689	6 BD107473	BD107473 Glucose-3
33	170	5.2	11332	1 AE011848	AE011848 Xanthomon
34	162	4.9	301214	1 AE016786	AE016786 Pseudomon
35	160	4.9	339681	1 AP003009	AP003009 Mesorhizo
36	159.5	4.9	2484	8 AY085885	AY085885 Arabidops
37	158.5	4.8	2247	8 ATH316231	ATH316231 Arabidops
38	158.5	4.8	2278	8 BT001942	BT001942 Arabidops
39	158.5	4.8	2385	8 AY039977	AY039977 Arabidops
40	158.5	4.8	43277	2 T5F17	T5F17 Arabidops
41	158.5	4.8	197655	8 ATCHRI69	ATCHRI69 Arabidops
42	157	4.8	10063	1 AE009366	AE009366 Agrobacte
43	157	4.8	10484	1 AE008247	AE008247 Agrobacte
44	152.5	4.6	13405	1 AE014492	AE014492 Brucella
45	149	4.5	332635	1 AP003005	AP003005 Mesorhizo

RESULT 1

ALIGNMENTS

BD013176 2106 bp DNA linear PAT 02-AUG-2002
LOCUS
DEFINITION A novel protein, a gene coding therefor and a method of using the same.
ACCESSION BD013176
VERSION BD013176.1 GI:22093365
KEYWORDS WO 0121657-A/1.
SOURCE Lyophyllum shimeji
ORGANISM Lyophyllum shimeji
Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes; Agaricales; Tricholomataceae; Lyophyllum.
REFERENCE 1 (bases 1 to 2106)
AUTHORS Takakura, Y., Kuwata, S. and Inoue, Y.
TITLE A novel protein, a gene coding, therefor and a method of using the patent: WO 0121657-A 1 29-MAR-2001;
JOURNAL JAPAN TOBACCO INC., CORPORATE JURIDICAL PERSON SOCIETY FOR TECHNO INNOVATION OF AGRICULTURE YUKIO SHIMOMURA, SHIRO TAKEKAWA, NOBIO CHO FORESTRY AND FISHERIES, YOSHIMITSU TAKAKURA, SHIGERU KUWATA, YASUHIRO INOUE
COMMENT OS Lyophyllum shimeji
PN WO 0121657-A/1
PD 29-MAR-2001
PE 20-SEP-2000 WO 2000JP006404
PR 21-SEP-1999 JP 99P 267238
PI YOSHIMITSU TAKAKURA, SHIGERU KUWATA, YASUHIRO INOUE PC
C07H14/375, C12N15/31, C12N15/63, C12N1/21, C12Q1/58, C12P21/02, PC A01N65/00
CC
FH Key Location/Qualifiers
FT CDS (8)..(1881).
FEATURES
source 1..2106
Location/Qualifiers
1..2106
/organism="Lyophyllum shimeji"
/mol_type="genomic DNA"
/db_xref="taxon:47721"
BASE COUNT 593 a 540 c 527 g 446 t
ORIGIN
Alignment Scores:
Pred. No.: 3,31e-247 Length: 2106
Score: 3284.00 Matches: 618
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: Gaps: 0
US-09-856-327-2 (1-618) x BD013176 (1-2106)
QY 1 MetSerLeuSerThrGluGlnMetLeuArgAspTyrProArgSerMetGlnIleAsnGly 20
DB 8 ATGTCTCTCTCAACGAGAGATGCTACCGCACTATCCACGGCTCATGCAAAATCAACGGA 67
QY 21 GlnIleProLysAsnAlaIleHisGlnThrTyrGlyAsnAspGlyValAspAlaPheIle 40
DB 68 CAGATTCTCTTAAGACGCAATTCACGAACATACGGAACGACGAGTTGATTTATTCATT 127
QY 41 AlaGlySerGlyProIleGlyAlaThrTyrAlaLysLeuGlyValGlnAlaGlyLeuArg 60
DB 128 GCAGATCTGCAGCCATTCGACGACGACGATGCAAAAGCTCTGTGGAAGCTGCTACGCT 187
QY 61 ValValMetValGlnIleGlyAlaAlaAspSerPheTyrAlaValAlaAsnAlaGlnGly 80
DB 188 GTTGTGATGCTGGAATCGGAGCTCTGTATCTTACGCTGTAAATGCCAAGAGAGA 247
QY 81 ThrAlaValProTyrValProGlyTyrHisLysLysAsnGlnIleGlnPheGlnLysAsp 100
DB 248 ACTGACAGTTCCTTACGTTCTCTGCTACCAACAAGATGAATGAGATTCACAGAAAGAT 307
QY 101 IleAspArgPheValAsnValIleLysGlyAlaLeuGlnAlaSerValProValArg 120
DB 308 ATTGACGCGTTCGCAATGATATCAAGAGACCTTACCAACAAGTCTCTGCTCTGCAGA 367
QY 121 AsnGlnAsnValProThrLeuAspProGlyAlaIlePheAlaProGlySerSerAla 140

|||||
DB 368 AACCAAGAACGCTGACACTTATCCGAGAGCTGGAGGCCGCCCTCGAAGTTCAGCC 427
QY 141 ILeSerAsnGlyLysAsnProHisGlnArgGlnPheGlnAsnLeuSerAlaGlnAlaVal 160
DB 428 ATATCGAACGCTGAAAAATCTCTACCAACGCGGAATTCGAGAACTTGTCTCGAGGCCGTA 487
QY 161 ThrArgGlyValGlyGlyMetSerThrHisTrpThrCysSerThrProArgIleHisPro 180
DB 488 ACGCGTGGATGGCGGATGAGTACCCACTGAGAGTGTCTCCACGCAAGATTCATCA 547
QY 181 PrometGlnSerLeuProGlyIleGlyAlaGlyProLysLeuSerAsnAspProAlaGlnAsp 200
DB 548 CCCATGGAAGTCTCCCGGACATCCGCTCCGAAGCTCAGTACGACACCCGGCAAGAGAC 607
QY 201 AspLysGlnTrpAsnGlnLeuTyrSerGlnAlaGlnArgIleGlyLysThrLys 220
DB 608 GACAAAGAGTGGACAGAGCTTATTCGAGGCCGAGCGCTCTATGCGGACTTCCACCAAG 667
QY 221 GluPheAspGlnSerIleArgHisThrLeuValLeuArgSerLeuGlnAspAlaTyrLys 240
DB 668 GAATTCGACGAGTCAATTGCGCACACCTTGTTCGGCTCTTTCGAAGACGCTACAAAG 727
QY 241 AspArgGlnArgIlePheArgProLeuProLeuAlaCysHisArgLeuLysAsnAlaPro 260
DB 728 GATCGTCAACGTAATCTTTCGCTCCCTCCGCTGGCATGCGACCGGTTGAAGACCGCGC 787
QY 261 GluTyrValGlnTrpHisSerIleAsnLeuAsnLeuPheHisSerIleTyrAsnAspLys 280
DB 788 GAATTCGCTGATGGCAGCTACAGCAAAATCTTTCCACTCTATCTACACAGATACAAAG 847
QY 281 GlnLysLysLeuPheThrLeuLeuThrAsnHisArgCysThrArgLeuAlaLeuThrGly 300
DB 848 CAGAAGAAGCTTATACCTGCTGAGCAACCATCTGCTGACACGACTGGCGCTACGGGC 907
QY 301 GlnTyrGlnLysLysIleGlyAlaAlaGlnValArgAsnLeuAlaThrArgAsnPro 320
DB 908 GGGTATGAGAAAGATATGGCGCTCCGCGAGGTACAGAACTTACTGGCCACGAAATCT 967
QY 321 SerSerGlnLeuAspSerTyrIleMetAlaLysValTyrValLeuAlaSerGlyAlaIle 340
DB 968 AGTTGCGACGCTGACAGCATATATCATGCGGAAGATATATGTACTGGCGTGGAGCGATC 1027
QY 341 GlnAsnProGlnIleLeuTyrAsnSerGlyPheSerGlyLeuGlnValThrProArgAsn 360
DB 1028 GGCAACCCACAGATTTCTATTAATCTGGGCTCTCGGCGTACAGTCCAGCCACGCAAT 1087
QY 361 AspSerLeuIleProAsnLeuGlnArgTyrIleThrGlnGlnProMetAlaPheGlyGln 380
DB 1088 GACTCGTTGATCCCAACCTGGGAGGTACATCAGCAGGACGACCGCATTTGTCAG 1147
QY 381 IleValIleArgGlnGluPheValAspSerValArgAspAspProTyrGlyLeuProTrp 400
DB 1148 ATAGCTTGAGGACAGAAATTCGTCAGACAGCGCGCGAGATCCTTATGAGACTGCCATG 1207
QY 401 TrpLysGlnAlaValAlaGlnHisIleAlaLysAsnProThrAspAlaLeuProIlePro 420
DB 1208 TGGAAAGAAAGCCGTTGCTCAACATATGCGCAAGAACCCGACAGATGACCTGCCATTC 1267
QY 421 PheArgAspProGlnProGlnValThrThrProPheThrGlnHisProThrPheHisThr 440
DB 1268 TTCGCGCATCCGAAACCCAGGTAACAACCCCATTTTACAGAAACACCCCTGGGACACG 1327
QY 441 GlnIleHisArgAspAlaPheSerTyrGlyAlaValAlaGlyProGlnValAspSerArgVal 460
DB 1328 CAGATTACCGCGGATGCTTTTTCGTACGCTGCGCTCGCTCGAGAGTGGACTCTGCTGTC 1387
QY 461 IleValAspLeuArgTyrPheGlyAlaThrAspProGlnAlaAsnAsnLeuValPhe 480
DB 1388 ATCGTCGACCTGCGCTGGTTGGCGCAACCGACCTGGAAGCAACACTTTTGGTTTTC 1447
QY 481 GlnAsnAspValGlnAspGlyTyrSerMetProGlnProThrPheArgTyrArgProSer 500
|||||

Db 1448 CAGACGATGTTCAAGCGGCTACAGTATGCCGACCGCGAGTTTCAGATATACGCCACG 1507
 Qy 501 ThrAlaSerAsnValaIarAlaArglysmetMetAlaAspMetCysGluValaIaSerAsn 520
 Db 1508 ACTGCGTCAACGTCAGACGCAAGAAATGATGGCCGATATGTGGCAATGGCGAGCAAC 1567
 Qy 521 LeuGlyGlyTyrLeuProThrSerProProGlnPheMetAspProGlyLeuAlaLeuHis 540
 Db 1568 TTGGAGAGTATATTGGCCACGACGCCCGCGAGTTATGATGATCCAGGCGCTTGCACTCAT 1627
 Qy 541 LeuAlaGlyThrThrArgIleGlyPheAspLysAlaThrThrValaAlaAspAsnSer 560
 Db 1628 CTTCGGGGGACTACTCGCATGGCTTGCACAAAGCAACATACAGTGGCTGATTAACACTTCG 1687
 Qy 561 LeuValTTPAspPheAlaAsnLeuThrValaAlaGlyAsnGlyThrIlaArgThrGlyPhe 580
 Db 1688 CTGGCTGGAGCTTGGCAATCTTATATGTTCGAGCAATGGCACCATTAGAGCGGCGCTTC 1747
 Qy 581 GlyLeuAsnProThrLeuThrSerMetCysHisAlaIleLysSerAlaArgSerIleIle 600
 Db 1748 GCGCGAAGCCGACACTTACGTCATGTGCCACGCTATCAAGAGCGCGAGCATCAATC 1807
 Qy 601 AsnThrLeuLysGlyGlyThrAspGlyLysAsnThrGlyGluHisArgAsnLeu 618
 Db 1808 AATACACTCAAGGGGTGGACTGACGCAAAATAATACAGCGAGCATCGCAACTT 1861
 RESULT 2
 AF535193
 LOCUS AF535193 2064 bp mRNA linear PLN 31-DEC-2002
 DEFINITION Peniophora sp. SG pyranose oxidase mRNA, complete cds.
 ACCESSION AF535193
 VERSION AF535193.1 GI:27436421
 KEYWORDS
 SOURCE Peniophora sp. SG
 ORGANISM Peniophora sp. SG
 Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes; Aphyllophorales; Lachnocladiaceae; Peniophora.
 REFERENCE 1 (bases 1 to 2064)
 Heekmann, D.M., Bastian, S., Rekowski, M., and Giffhorn, F.
 Pyranose Oxidase of the White-Rot Fungus Peniophora sp. Strain SG: Cloning and Characterization of the Gene, Heterologous Expression in *Escherichia coli*, and Properties of the Recombinant Enzyme Unpublished
 2 (bases 1 to 2064)
 Heekmann, D.M. and Giffhorn, F.
 Direct Submission
 JOURNAL Submitted (07-AUG-2002) Applied Microbiology, University of the Saarland, Im Stadtwald, 66123 Saarbruecken, Saarland, Germany
 FEATURES
 source
 1..2064
 /organism="Peniophora sp. SG"
 /mol_type="mRNA"
 /strain="SG"
 /db_xref="taxon:204723"
 26..1897
 /EC_number="1.1.3.10"
 /note="FAD-oxidoreductase; Pox; pyranose:oxygen
 2-oxidoreductase; poxSG"
 /codon_start=1
 /product="pyranose oxidase"
 /protein_id="AA013382.1"
 /db_xref="GI:27436422"
 /translation="MSTSSDPFNFNAKSPFSAKAAKASASLPLRPGPKKVGMD
 IKYDVVIGSGPCTCYARELYAGYKRAMDGEIDSGIKGAKKTVYKKNIDK
 FVNYIOGLMSVSPVNTLVDTLSPTSMQSTFVRNGSPEDDPLTSGOAVTRY
 VGMSTHTCATPFRDREPLVLKDDADADAEDRLYTAESYFQGTQKESIR
 HNLVNLKAEYKQDFQOQIPLATRSPFVWSSANTVFEDONRNTDAPEREN
 LEPVAVKERVYRNALNREISLHIDLSGRFEIKAVYVLTGAINTDOLVSGF
 GOLGRPNTPNPELLSLGSIYTESOILFECOTVSTELIDVKSOMTIRGNGELTS
 VTPVPGASTNKHPMMNKKVKNHMOHOEDLPIPEPPEOVYTLTPSRPHQIH
 RDAFSYGAVOOSIDSRLLVDMRFGRTRPKENKIMPSDKTTDAYNNPQPTDFRPA
 GRISKEADMTDMCVMSAKIGFLPGSLPQPMEGVLHLHGTHRMKDFDEKNCV
 NTDSRVGFEKNIPLGGCGNIPITAYGANTPLAMSIAIKSCEYIKONFTPSPTSBAQ*

polyA_site 2047
 BASE COUNT 462 a 652 c 579 g 371 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 2,32e-87
 Score: 1299.50
 Percent Similarity: 57.36%
 Best Local Similarity: 43.29%
 Query Match: 37.74%
 DB: 8 Gaps: 16
 US-09-856-327-2 (1-618) x AF535193 (1-2064)
 Qy 37 AspValPheIleAlaGlySerGlyProIleGlyAlaThrTyrAlaLysLeuGluValGlu 56
 Db 167 GACGTTGCATATGATGAGCTCCGACCGATGTGATGACAGTATGCCGCGAGCTGCGC 226
 Qy 57 AlaGlyLeuArgValAlaMetValGluIleGlyAlaAlaAspSerPheTyrAlaValAsn 76
 Db 227 GCAGGTTACAGAGTCGCATGTGCACATCGAGAGATCGACTCGGCTGAAGTT--- 283
 Qy 77 AlaGluGluGlyThrAlaValProTyrValProGlyTyrHisLysLysAsnGluIleGlu 96
 Db 284 -----GGTCCCAACAGAAACACACCTCGAG 310
 Qy 97 PheGlnLysAspIleAspArgPheValAsnValIleLysGlyAlaLeuGlnGluValSer 116
 Db 311 TACCAAGAACATCGACAGAGTTGTGATGACGATTCAGGGCAATTAATGCTGTTTCC 370
 Qy 117 ValProValArgAsnGlnAsnValProThrLeuAspProGlyAlaIleArgSerAlaProPro 136
 Db 371 GTTCCCGCATATACCTCGTGTGTCACACGCTACCCGAGCTGTTGGCAAGCT----- 424
 Qy 137 GlySerSerAlaIleSerAsnGlyLysAsnProHisGlnArgGluPheGluAsnLeuSer 156
 Db 425 TCGACGTTCTTCTCCGACGCGCTCGAAGCCAGACGCCCTTCGTAACTCAGT 484
 Qy 157 AlaGluAlaValThrArgGlyValGlyLysMetSerThrHisThrThrCysSerThrPro 176
 Db 485 GGTCAAGGGGTACCGCGGTGCTCGAGACATGTACGACATGGACATGGCCACACCC 544
 Qy 177 ArgIleHisProProMetGluSerLeuProGlyIleGlyArgProLysLeuSerAsnAsp 196
 Db 545 CGTTTCGACCGGACGACAA-----CGCGCGCTGCTGTGAAGAC 583
 Qy 197 ProAlaGlu---AspAspLysGluThrAsnGluLeuTyrSerGluAlaGluArgLeuIle 215
 Db 584 GACGCGGACGCTGACGACGCTGAATGGACCGCTACACCAAGCGGACGCTGTAATTC 643
 Qy 216 GlyThrSerThrLysGluPheAspGluSerIleArgHisThrLeuValLeuArgSerLeu 235
 Db 644 CAGACTGGACGACGACGCTCAAGGATGATCCGCCAACACTCTGCTCAACAGCTC 703
 Qy 236 GlnAspAlaTyrLysAspArgGlnArgIlePheArgProLeuProLeuAlaCysHisArg 255
 Db 704 GCGAGGAAATACAAAGGC---CAACGCACTTCACAGAGATTCACCTGCGCGGACCGCC 760
 Qy 256 LeuLysAsnAlaProGluTyrValGluThrPheSerAlaGluAsnLeuPhe-----His 273
 Db 761 -----CGAGCCCGACCTGCTGGAATGAGACTGGGAAACACAGTGTTCAGCTCCAG 814
 Qy 274 SerIleTyrAsnAspAspLysGluLysLysLeuPheThrLeuLeuThrAsnHisArgLys 293
 Db 815 AACAGCCGACACGAGGAGCGCGGAGGAGGCTTCAACCTTCCCGCGGTCGCGGT 874
 Qy 294 ThrArgLeuAlaLeuThrGlyGlyTyrGluLysLysIleGlyAlaAlaGluValArgAsn 313
 Db 875 GAGCGCGTC-----GTGGGCAAC 892
 Qy 314 LeuLeuAlaThrArgAsnProSerSerGlnLeuAspSerTyr----- 327
 Db 893 GCGTTG-----AACTCGGAGATCGAGAGTCTGCACATCCAGACCTCAATC 937

```
QY 328 -----IleMetAlaLysValTyrValLeuAlaSerGlyAlaIleGly 341
Db 938 TCGGCGGACGGCTTCGAAATCAAGGCTGACGTGCTCTCCACCGGGGGGCTGCAC 997
QY 342 AspProGlnIleLeuTyrAsnSerGlyPheSerGlyLeu---GlnValIleProAlaGan 360
Db 998 AACACGCGACTTCTGTAATCTGGCTTGGACACGCTGGCGCGAACCACCAAC 1057
QY 361 ---AspSerLeuIleProAlaSerLeuGlyAArgTyrIleThrGluIleProMetAlaPheCys 379
Db 1058 CCACCGGAGCTGTGTCGCCGCCCTCGGGGAGCTACATACCGAAGAGTCGCTCTCTGC 1117
QY 380 GlnIleValLeuArgGlnIlePheValAlaSerValAlaAspAspPro-----TyrGly 397
Db 1118 CAGACCGTGATGAGCACCAGCATCATGACAGCGTCAAGTCCGATCCGATCAGGGGA 1177
QY 398 LeuPro----- 399
Db 1178 ACCCTGGCGAGCTGACGTACAGCGTACAGCCGCGCGCTGCACCAACAGCAC 1237
QY 400 -----TriPTrpLysGluAlaValAlaGlnHisIleAlaLysAsnProThrAspAlaLeu 417
Db 1238 CCGGACTGGTGGACAGCAAGGTGAATAAACACATGATGACGACCAAGAGACCCGCTC 1297
QY 418 ProIleProPheArgAspProGluProGlnValIleThrProPheThrGluGlnHisPro 437
Db 1298 CCGATCCGCTGAGAGACCCCGAGCGGAGGTACACACTGTTCCACCGCTGCACCCG 1357
QY 438 TrpHisThrGlnIleHisArgAspAlaPheSerTyrGlyAlaValGlyProGluValAsp 457
Db 1358 TGGCACACTGATCCACCGCATGTCTTACGTACGGCGCATGGCAGCAAAACAGCAGC 1417
QY 458 SerArgValIleValAspLeuArgTyrPheGlyAlaThrAspProGluAlaAsnAsnLeu 477
Db 1418 TCGCGCTCATGTGGATGGCGCTCTTCGGCGGAGCGGAGCCCAAGAGGAGCAAG 1477
QY 478 LeuValPheGlnAsnAspValGlnAspGlyTyrSerMetProGlnProThrPheArgTyr 497
Db 1478 CTCTGGTCTCTCAGACAGATTCACGGCGCTCAACATCCGACGCCACGTTGCGACTTC 1537
QY 498 Arg---ProSerThrAlaSerAsnValAlaArgIleGlyMetMetAlaAspMetGlySer 516
Db 1538 CCGTTCGCCGGCGCGCGCGCACAGCAAGGCGGAGACATGATGACCGCATGTGCTT 1597
QY 517 ValAlaSerAsnLeuGlyTyrLeuProThrSerProProGlnPheMetAspProGly 536
Db 1598 ATGTGCGGAGATGTGGTGGCTTCTTACCCTGCTCCCGCAATTCAATGAGCCTGCT 1657
QY 537 LeuAlaLeuHisLeuAlaGlyThrThrArgIleGlyPheAspIleAla-----ThrThr 554
Db 1658 CTTGTCCTTCCACCTTGGTGTACGCGACCCCATGGGCTTGCAGAGAGAGCAACTGC 1717
QY 555 ValAlaAspAsnSerLeuValIleThrAspPheAlaAsnLeuTyrValAlaGlyAsnGly 574
Db 1718 TCGGTCAACACGAGACTCGCGCGTTCGCTTCAGAACCTCTCTCTCGTGGCTGGGA 1777
QY 575 ThrIleArgThrGlyPheGlyGlnAsnProThrLeuThrSerMetCysHisAlaIleLys 594
Db 1778 AACATTCTTACCGCGATGCGGCGAAGCCGACGCTACCGCAATGTGCTCGCGATCAAG 1837
QY 595 SerAlaArgSerIle 599
Db 1838 AGTTGCGAGTACATC 1852
RESULT 3
LOCUS AY291124 2069 bp mRNA linear PLN 23-MAY-2003
DEFINITION Trimetes ochracea strain MB49 pyranose oxidase (p2o) mRNA, complete
cds.
ACCESSION AY291124
VERSION AY291124.1 GI:31044223
KEYWORDS
```

```
SOURCE
ORGANISM Trimetes ochracea
Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
Aphyllophorales; Trimetes.
REFERENCE
AUTHORS Vecerek, B., Maresova, H., Kocanova, M. and Kyslik, P.
TITLE Molecular cloning and expression of the pyranose-2 oxidase gene
from Trimetes ochracea MB49 in Escherichia coli
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2069)
AUTHORS Vecerek, B., Maresova, H. and Kyslik, P.
TITLE Direct Submission
JOURNAL Submitted (06-MAY-2003) Institute of Microbiology AS CR, Videnska
1083, Prague 4 142 20, Czech Republic
FEATURES
source
1..2069
location/Qualifiers
1..2069
/organism="Trimetes ochracea"
/mol_type="mRNA"
/strain="MB49"
/db_xref="taxon:230624"
1..2069
/gene="p2o"
49..1920
/codon_start=1
/product="pyranose oxidase"
/protein_id="AAP40332.1"
/db_xref="GI:31044224"
/translation="MSTSSDPFPNFKSSPSRAAOKASASSLPPLPGDPKKVGM
IKDYIVYSSGPIGCTYARELVAGCYKVAEMDIEGIDSLKLGAKKKNVEOKNIDK
FVNVIQGLMSVSVFVNITLVDTLSTMSQASTFVRNNSNEQDPLRNLSQAVTR
VGMSTHWTCATPREDEQRPPLIVADDAEDALAEWRLYTKAESYFQRTGDFKESIR
HNLVLNKLIEEYKQGRDFOQIPLAATRRSPTEVWSSANTVPDQLQRPNTDAPBEFN
LPPAVACERVVKNALNIESLHIDILISGDRFEIKADYVLTAGVHTQLVSGF
GOLGRPNPAMPPELLVLSGSLYTESVLCQVPMSEILDSVSKDWTIGTGELTYS
VYVTGASNKRHPDMNKEVKNHMMOHQDPLIPREDEPOVTLTFQSPHMHQIH
RDAPSGAVQOSIDSLRLYDNRPFRTKEKKNLIFSKTIDATNMPPTDFRPA
GRTSKAEEDMTDMCVMSAKTIGFLPGLSPQWPEPLVLTGTHRMGDEKEDNCV
NDSRVFEGKRLFLGCGNIPRAYGANPTLTAMSLAIKSCVEYIKQNFPSPTSEAQ"
BASE COUNT 449 a 659 c 584 g 377 t
ORIGIN
Alignment Scores:
Pred. No.: 2,33e-87 Length: 2069
Score: 1239.50 Matches: 274
Percent Similarity: 57.36 Conservative: 73
Best Local Similarity: 45.29 Mismatches: 173
Query Match: 37.74 Indels: 85
DB: 8 Gaps: 16
US-09-856-327-2 (1-618) x AY291124 (1-2069)
QY 37 AspValPheIleAlaGlySerGlyProIleGlyAlaThrTyrAlaLysLeuGlyVal 56
Db 190 GACGTTGTATGTAGAGCTCCGACCGATTTGATGACGACCTAATCCCGTAGCTCGTCCG 249
QY 57 AlaGlyLeuArgValAlaMetValGluIleGlyAlaAlaAspSerPheTyrAlaValAsn 76
Db 250 GCAGGTTCACAGGTGCCCATGTTGCACATCGGAGAGATCGACTCTGGCCTGAAGATT 306
QY 77 AlaGluGluGlyThrAlaValProTyrValProGlyTyrHisIleLysAsnGluIleGlu 96
Db 307 -----GGTGGCCCAACAGAAACACCGTCGAG 333
QY 97 PheGlnLysAspIleAspArgPheValAsnValIleLysGlyAlaLeuGlnIleValSer 116
Db 334 TACCAAGAGACATGACAGATTGTGACGATTCATGAGGCAACTGATGTCTGTTTCC 393
QY 117 ValProValArgAsnGlnAsnValProThrLeuAspProGlyAlaTyrSerAlaProPro 136
Db 394 GTTCCCGTCATACCTCGTGTGTCGACACGTCACGCCGACGCTTGGCAAGCT----- 447
QY 137 GlySerSerAlaIleSerAsnGlyLysAsnProHisGlnArgGluPheGlnAsnLeuSer 156
```



```

Db      448 TCGACGCTTCCTGCGCAACGCTCGAACCCAGAGCCACCTTCGTAACTCCAGT 507
Qy      157 AlaGluValAlaThrArgGlyValGlyMetSerThrHisThrPheSerThrPro 176
Db      508 GGTGAGGGGTCAGCGCGTCTGCGAGGCAATGTCATCGACATGGCCACACC 567
Qy      177 ArgIleHisProMetGluSerLeuProGlyIleGlyArgProLysLeuSerAsnAsp 196
Db      568 CGTTTCAGACCGCGAGCAA-----CGCCCGCTCTCTGTGAGAGAC 606
Qy      197 ProAlaGlu---AspAspLysGluTrpAsnGluLeuTyrSerGluAlaGluArgLeuIle 215
Db      607 GACCGGAGCGCTGACGAGCTGAGTGGACCGCTCTACACCAAGCGGAGCTGACTTC 666
Qy      216 GlyThrSerThrLysGluPheAspGluSerIleArgHisThrIleValLeuArgSerLeu 235
Db      667 CAGACTGGAGCGACAGCTTCAGAGAGTCGATCCGCCAATCTGCTCTCAACAGCTC 726
Qy      236 GluAspAlaTyrLysAspArgGlnArgIlePheArgProLeuProLeuAlaCysHisArg 255
Db      727 ACGGAGGAATCAAGGGC---CAACGCGACTTCCAGCAGATTCCACTCGCGCAACCCGC 783
Qy      256 LeuLysAsnAlaProGluTyrValGluTrpHisSerAlaGluAsnLeuPhe-----His 273
Db      784 -----CGAGCCCGACCTCTGCGAATGAGCTCGCGGACACACTGTTTTCGACTCCAG 837
Qy      274 SerIleTyrAsnAspAspLysGluLysLysLeuPheThrLeuLeuThrAsnHisArgCys 293
Db      838 AACAGGCGGAGACGAGCGCGCGAGAGCGCTTCAACCTTCCCGCGGTGCGCGT 897
Qy      294 ThrArgLeuAlaLeuThrGlyGlyTyrGlyLysLysIleGlyAlaAlaGluValArgAsn 313
Db      898 GAGGCGCTC-----GTGCGCAAC 915
Qy      314 LeuLeuAlaThrArgAsnProSerSerGlnLeuAspSerTyr----- 327
Db      916 GCGTTG-----AACTCGGAGATCGAGATCTGCACATCCAGACCTCATC 960
Qy      328 -----IleMetAlaLysValTyrValIleLeuAlaSerGlyAlaIleGly 341
Db      961 TCGGGCGACCGCTTCGAATCAAGCTGACCTGTAACGCTCACCGCGGGCGCTCCAC 1020
Qy      342 AsnProGlnIleLeuTyrAsnSerGlyPheSerGlyLeu---GlnValThrProArgAsn 360
Db      1021 AACACGCGACTTCTCGTGAACCTGCGCTTGGACAGCTGGCGCCGACCAACCCGCAAC 1080
Qy      361 ---AspSerLeuIleProAsnLeuGlyArgTyrIleThrGlnGlnProMetAlaPheCys 379
Db      1081 CCACCGGAGCTGCGCGTCCCTGGGAGACTACATCACCGACGAGCAGTCTGCTCTGCG 1140
Qy      380 GlnIleValIleLeuArgGlnGluPheValAspSerValArgAspAspPro-----TyrGly 397
Db      1141 CAGACCGGTGATGACACCGACTCATCGACGCTCAAGTCCGACATGACCATCAGGGGA 1200
Qy      398 LeuPro----- 399
Db      1201 ACCCTTGGCGGAGCTGACGTACAGCGTACACGCCGCGCGCTGCACCAACAGACAC 1260
Qy      400 -----TrpTyrLysGluAlaValAlaGlnHisIleAlaLysAsnProThrAspAlaLeu 417
Db      1261 CCGGACTGTGTAAGCAAGAGTGAATAAACACATGATGACAGCACAGAGAGACCGGCTC 1320
Qy      418 ProIleProPheArgAspProGluProGlnValThrThrProPheThrGluGluHisPro 437
Db      1321 CCGATCCCGTTCGAGGCCCGAGCGCGAGTTCACCTCTGTTCAACCGTGCACCGC 1380
Qy      438 TrpHisThrGlnIleHisArgAspAlaPheSerTyrGlyAlaValGlyProGluValAsp 457
Db      1381 TGGGACACTGATCCACGCGGATGCTTTCAGTTACGGCGGACATGACCAAGCATGCAC 1440
Qy      458 SerArgValIleValAspLeuArgTrpPheGlyAlaThrAspProGluAlaAsnLeu 477

```

```

Db      1441 TCGCGTCTCATCGTAGCTGGCGCTTCTTGGCGCGAGCGAGCCCAAGAGAGAACAG 1500
Qy      478 LeuValPheGluAsnAspValGluAspGlyTyrSerMetProGlnProThrPheArgTyr 497
Db      1501 CTCTGGTTCTCAGACAAATCATCAGCAGCCGTCATCAACATGCCGACGCCACGCTTGACTTC 1560
Qy      498 Arg---ProSerThrAlaSerAsnValArgAlaArgLysMetMetAlaAspMetCysGlu 516
Db      1561 CGCTTCCCGCGCGCGCGCGAGCAGAGAGCGGAGCAGCATGATGACCGACATGTCGCTT 1620
Qy      517 ValAlaSerAsnLeuGlyGlyTyrLeuProThrSerProProGlnPheMetAspProGly 536
Db      1621 ATGTGCGCGAAGATCGGTGGCTTCTTACCCGCTCCCTCCCGCAATTCATGAGCTGTGT 1680
Qy      537 LeuAlaLeuHisIleAlaGlyThrThrArgTyrIleGlyPheAspLysAla-----ThrThr 554
Db      1681 CTGTGCTTTCACCTTGTGTGTGATGACCGCCGATGGGCTTCGACGAGAGAGACACTGC 1740
Qy      555 ValAlaAspAsnAsnSerLeuValTrpAspPheAlaAsnLeuTyrValAlaGlyAsnGly 574
Db      1741 TGGCTCACACGAGACTCGCGGTGTGTGCGTTCAAGAACCTTCTCGTGGTGGCGGA 1800
Qy      575 ThrIleArgThrGlyPheGlyGluAsnProThrLeuThrSerMetCysHisAlaLys 594
Db      1801 AACATCTCTACCGGTACGCGCGCGAGACCGAGCTCACCGGAAATGTCGTCGATCAAG 1860
Qy      595 SerAlaArgSerIle 599
Db      1861 AGTTGCGAGTACATC 1875

RESULT 4
LOCUS      E11766                      1869 bp    RNA      linear    PAT 29-SEP-1997
DEFINITION E11766 of pyranose oxidase.
ACCESSION  E11766.1 GI:22025402
VERSION     JP 1996205861-A/1.
KEYWORDS   pyranose oxidase
SOURCE      unclassified
ORGANISM   unclassified.
            1 (bases 1 to 1869)
REFERENCE   1 (bases 1 to 1869)
            Nishimura,I., Okada,K., Nanbara,T., Kawai,G., Koyama,T. and
            Suzuki,M.
            NEW PYRANOSE OXIDASE, PYRANOSE OXIDASE GENE, NEW RECOMBINANT DNA
            AND PRODUCTION OF PYRANOSE OXIDASE
            Patent: JP 1996205861-A 1 13-AUG-1996;
            KIRKMAN CORP
            OS      Coriolus versicolor
            PN      JP 1996205861-A/1
            PD      13-AUG-1996
            PF      24-MAY-1995 JP 1995124835
            PR      07-DEC-1994 JP 94P 304086
            PI      NISHIMURA IKUKO, OKADA KIMI HARU, NANBARA TOMOYUKI, PI
            GENSHIRO,
            KOYAMA TAIJI, SUZUKI MASARU
            PC      C12N9/04,C12N15/09,(C12N9/04,C12R1.19),(C12N15/09,C12R1.01);
            CC      strandedness: Double;
            FH      key
            FT      source 1.1869
            /strain='ps4a',
            /location/Qualifiers
            1.1869
            /organism='unidentified'
            /mol_type='genomic RNA'
            /db_xref='taxon:32644'
BASE COUNT  408 a 593 c 529 g 339 t
ORIGIN
Alignment Scores: 7.12e-87 Length: 1869
Pred. No.:

```

Score:	1232.50	Matches:	273
Percent Similarity:	57.00%	Conservative:	73
Best Local Similarity:	44.98%	Mismatches:	176
Query Match:	37.53%	Indels:	85
	6	Gaps:	16
US-09-856-327-2 (1-618) x E11766 (1-1869)			
QY	37	AspValPhe11eAlaGlySerGlyPro11eGlyAlaThrTyralaLysLeuGlyValGlu	56
	142	GAGCTTGTCATAGAGGCTCCGGACCGATTGATGACGATGATGCGCTGAGCTGCGAA	201
QY	57	AlaGlyLeuArgValValMetValGlu11eGlyAlaAlaAspSerPheTyralaValAsn	76
	202	GCGGCTTACAAAGGTCGCGCATGTCGACATCGGGAAATTAATCTTGCGCTGAGATC---	258
QY	77	AlaGluGluGlyThrAlaValProTyValProGlyTyThrIAspLysLysAsnGlu11eGlu	96
	259	-----GCTCCCAAGAACAGAACACCGCTCGAA	285
QY	97	PheGlnLysAsp11eAspArgPheValAsnVal11eLysGlyAlaLeuGlnValSer	116
	286	TACACAGAAAGACATTGACAAATTGTGTAACGTATTCAGGCGCAATGATGCTGTTCC	345
QY	117	ValProValArgAsnGlnAsnValProThrLeuAspProGlyAlaTrpSerAlaProPro	136
	346	GTTCCCGCTCAATACCTCGTGCATCGACACGCTCAGCCCGCTTGCGCAAGCT----	399
QY	137	GlySerSerAla11eSerAsnGlyLysAsnProHisGlnArgGluPheGlnAsnLeuSer	156
	400	TCACTGCTTCGTCGCCATGGGTGCGAACCCAGACAGAGACCGCTTCTTAACCTCAGT	459
QY	157	AlaGluAlaValThrArgGlyValGlyLysSerThrHisThrCysSerThrPro	176
	460	GGTGAGCGGTCACGCTGTGTCGAGGCGATGTCACGACGATGATGCGGACACCG	519
QY	177	Arg11eHisProPheMetGluSerLeuProGly11eLysArgProLysLeuSerAsnAsp	196
	520	CGCTTTGACCGCAGCAG-----CGCCCGTGTGCTGTAAGAC	558
QY	197	ProAlaGlu---AspAspLysGluTrpAsnGluLeuTyrsersGluAlaGluArgLeu11e	215
	559	GACGAGAGCTGACGCGCGAGTGGGACCGGCTGACACCAAGCGCGAGTATCTTC	618
QY	216	GlyThrSerThrLysGluPheAspGluSer11eArgHisThrLeuValLeuArgSerLeu	235
	619	AAGACCGGAGCAGATTCAAGAGTGCATCCGCGACAACTCGCTGCAACAAGCTC	678
QY	236	GlnAspAlaTyrtLysAspArgGlnArg11ePheArgProLeuAlaCysHisArg	255
	679	GCGGAGGAATACAAAGT---CAGCGCGACTTCACAGACATCCCGCTCGCGGACGCGT	735
QY	256	LeuLysAsnAlaProGluTyValGluTrpHisSerAlaGluAsnLeuPhe-----His	273
	736	-----CGCAGTCGACCTTCGTGAGTGGAGCTCGCGAACAACCGTTCGACCTCCAG	789
QY	274	Ser11eTyAsnAspLysGlnLysLeuPheThrLeuLeuThrAsnHisArgCys	293
	790	AACAGCGCGAACAACGAGCGCGCAATGAGCGCTTCACCTCTCCCGCGGCTGCATGT	849
QY	294	ThrArgLeuAlaLeuThrGlyTyGlyTyGluLys11eGlyAlaAlaGluValArgAsn	313
	850	GAGCGGCTC-----GTGCGCAAC	867
QY	314	LeuLeuAlaThrArgAsnProSerSerGlnLeuAspSerTyR-----	327
	868	-----ACGTGCAACTCCGAGATCGAGATCTGCACATCCACGACCTCATC	912
QY	328	-----11eMetAlaLysValTyValLeuAlaSerGlyAla11eGly	341
	913	TGGGGGACCGCTTCGAAATCAAAAGACGAGCGTGTCTTACACGCGGCGGCTCCAC	972
QY	342	AsnProGln11eLeuTyAsnSerGlyPheSerGlyLeu---GlnValThrProArgAsn	360

Db	973	AACGGGAGCTTCTCGTGAACCTCGCTTGGACACGTGGCGCGGACCCCGCAAC	1032
QY	361	---AspSerLeu11eProAsnLeuGlyArgTyThrGlnGluInProMetAlaPheCys	379
	1033	CCGCGGAGTTCGCTCGCGCTCCCTGGAGGTACATCCACGACAGTGGCTTCTTCC	1092
QY	380	Gln11eValLeuArgGlnGluPheValAspSerValArgAspAspProTyR-----Gly	397
	1093	CAGACCGTATGAGACACCGAGCTCATGACAGCGTCAAGTCCGACATGATCATCAGGGG	1152
QY	398	LeuPro-----	399
	1153	AACCTGCGCATGTGGGTACACGCTCACGACGCGCGCGGACCAACCAAGCAC	1212
QY	400	-----TrpTrpLysGluAlaValAlaGlnHis11eAlaLysAsnProThrAspAlaLeu	417
	1213	CCGAGCTGTGGAACGAAAGTGAAGAACACATGATGACGACCAAGAGACCCGCTT	1272
QY	418	Pro11eProPheArgAspProGluProGlnValThrThrProPheThrGlnHisPro	437
	1273	CCAATCCCGTTCGAGGACCCCGAGCCGACAGTCAACCTTTCAGCCATCGCACCCG	1332
QY	438	TrpHisThrGln11eHisArgAspAlaPheSerTyGlyAlaValGlyProGluValAsp	457
	1333	TGGCACACTGATTCACCGCGATGCTTCACTTACGTCGCGCGGTGACCAAGCATCGAC	1392
QY	458	SerArgVal11eValAspLeuArgTrpPheGlyAla11eAspProGluAlaAsnLeu	477
	1393	TCACTGCTCATCTCGACATGCGCTTCTTGCGCGGAGCGGACCAAGAGGACCAAG	1452
QY	478	LeuValPheGlnAsnAspValGlnAspGlyTyRSerMetProGlnProThrPheArgTyR	497
	1453	CTCTGCTTCGCGCAAAATTAACGACACATACATGCGCGACGCGAGCTTGCACCTTC	1512
QY	498	Arg---ProSerThrLysSerAsnValArgAlaArgLysMetMetAlaAspMetCysGlu	516
	1513	CGCTTCCCGGCGCGGCGCACGACCAAGAGCGGAGACATGATGACCGATATGTGCGTT	1572
QY	517	ValAlaSerAsnLeuGlyTyRLeuProThrSerProProGlnPheMetAspProGly	536
	1573	ATGTCGGCAAGATGTGTGCTTCGTGCGCGGCTCCCTCCGCAATTCATGAGACCCGT	1632
QY	537	LeuAlaLeuHisLeuAlaGlyThrThrArg11eGlyPheAspLysAla-----ThrThr	554
	1633	CTTGCTCTTCACTCGCTGCTGATGACGACCGATGCGCTTCAGCAGGACGAGGACAACTGC	1692
QY	555	ValAlaAspAsnAsnSerLeuValTrpAspPheAlaAsnLeuTyValAlaGlyAsnGly	574
	1693	TGCGTCAACACGAGCTCGCGCTGTGGCTTCAAGAACCTGTCTCGGTGGCTCGGA	1752
QY	575	Thr11eArgThrGlyPheGlyGluAsnProThrLeuThrSerMetCysHisAla11eLys	594
	1753	AACATTCGCCACCGGACGCGGCAACCCGAGCTCACCGCATGTGCGTGGATCAAG	1812
QY	595	SerAlaArgSer11eLeasn	601
	1813	AGTTGCGAGTACATCAAGAAC	1833
RESULT 5			
LOCUS	182410	1869 bp	DNA
DEFINITION	Sequence 1 from patent US 5712139.		
ACCESSION	182410		
VERSION	182410.1	GI:3210707	
KEYWORDS	Unknown.		
SOURCE	Unknown.		
ORGANISM	Unclassified.		
REFERENCE	1 (bases 1 to 1869)		
AUTHORS	Nishimura,T., Okada,K., Minamihara,T., Kawai,G., Koyama,Y. and Suzuki,M.		
TITLE	Pyranose oxidase, pyranose oxidase gene, novel recombinant DNA and		

process for producing pyranose oxidase
JOURNAL Patent: US 5712139-A 1 27-JAN-1998;
FEATURES Location/Qualifiers
source 1..1869
/organism="unknown"

BASE COUNT 408 a 593 c 529 g 339 t
ORIGIN

Alignment Scores:

Pred. No.:	7 12e-87	Length:	1869
Score:	1232.50	Matches:	273
Percent Similarity:	57.00%	Conservative:	73
Best Local Similarity:	44.98%	Mismatches:	176
Query Match:	37.53%	Indels:	85
DB:	6	Gaps:	16

US-09-856-327-2 (1-618) x 182410 (1-1869)

```

QY 37 AspValIphelelaaglyserglyproileglyalathrTyralAlaLysLeuCyValGlu 56
    |||||  |||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 142 GACGTGTTCATAGTAGGCTCCGACCGATGTGATGCACGTATGCCCGTACGTCTCGAA 201
QY 57 AlaGlyLeuArgValIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 76
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 202 GCCGGTTACAAAGCTGCCATGTTCCACATCGGGAAATTAATGCTGCGCTGAAGATC--- 258
QY 77 AlaGluGluGlyThrAlaValIProGlyValIProGlyTyRHisLysLysAsnGluIleGlu 96
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 259 -----GGTGGCCCAAGAGAACACCCCTCGAA 285
QY 97 PheGlnLysAspIleAspArgPheValAsnValIleLysGlyAlaLeuGlnGlnValSer 116
    ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
DB 286 TACAGAGAACATGATGACAGATTGTGAACGTCATTCAGGGCCCAATGATGTCGTTC 345
QY 117 ValIProValArgAsnGlnAsnValIProThrIleuAspProGlyAlaItrPserAlaProPro 136
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 346 GTTCCCGTCATACCTCGTATCGACACGACCGACCGACCTGTGGCAACCT----- 399
QY 137 GlySerSerAlaIleSerAsnGlyLysAsnProHisGlnArgGluPheGlnAsnLeuSer 156
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 400 TCATGCTCTTCGTCGCAATGGCTCGAATCCAGACGAGACCCGCTTCGTAACCTCACT 459
QY 157 AlaGlnAlaValIThrArgGlyValGlyLysMetSerThrHisTrpThrCysSerThrPro 176
    ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
DB 460 GGTCAAGGGGTCACCGCTGTCGTGAGGACATGTCACGACATGACATGCGCCAGACCG 519
QY 177 ArgIleHisProIleMetGluSerLeuProGlyIleGlyArgProLysLeuSerAsnsp 196
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 520 CGCTTTGACCGGACAG-----CGCCCTTGTCTGTAAGGAC 558
QY 197 ProAlaGlu-----AspAspLysGluItrPAsnGluLeuTyRserGluAlaGluArgLeuIle 215
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 559 GACCAAGACGCTGACGAGCGCGAGTGGACCGGCTGTACACCAAGGCGGAGTCATCTTC 618
QY 216 GlyThrSerThrLysGluPheAspLysIleArgHisThrLeuValLeuArgSerLeu 235
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 619 AAGACGGGACGAGCAGTCAAGAGATCGATCCGCCAACACCTGCTGTCAACAACGTC 678
QY 236 GlnAspAlaItyrLysAspArgGlnArgIlePheArgProLeuProLeuAlaCysHisArg 255
    ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
DB 679 GCGGAGGAATCAAAAGT---CAGCGCAACCTTCACGAGATCCCTCGCGGCAACGCGT 735
QY 256 LeuLysAsnAlaProGluTyRValGluItrPHisSerAlaGluAsnLeuPhe-----His 273
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 736 -----CGCAGTCCACCTTCGTCAGTGGAGCTGGCGAACACCGTGTTCACCTCCAG 789
QY 274 SerIleTyRAsnAspLysGlnLysLysLeuPheThrIleuLeuThrAsnHisArgCys 293
    ::|||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 790 AACAGCGGACGACGAGCGCGGAGTGAAGCGCTTCAACCTCTTCGCCGCGTGCATGT 849
QY 294 ThrArgLeuAlaLeuThrGlyGlyTyRgluLysLysIleGlyAlaAlaGluValArgAsn 313
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 850 GAGCGCGTC-----GTGGCGAAC 867

```

```

QY 314 LeuLeuAlaThrArgAsnProSerSerGlnLeuAspSerTyR----- 327
DB 868 -----ACGTGCAACTCCGAGATCGAGAGCTTCACATCCACGACCTCATC 912
QY 328 -----IleMetAlaLysValTyRValLeuAlaSerGlyAlaIleGly 341
DB 913 TCGGGCAGACCGCTTCGAATTAACAGCAGAGCTGTTCGTTCTTAACAGCGGGCGGTCCAC 972
QY 342 AsnProGlnIleLeuTyRAsnSerGlyPheSerGlyLeu---GlnValThrProArgAsn 360
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 973 AACGGCAGACCTTCGTGAACTCTGCTTGGACAGCTGGCGCGGACCGGACCGCGAAC 1032
QY 361 ---AspSerLeuIleProAsnLeuGlyArgTyRThrGluGlnIleProMetAlaPheCys 379
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 1033 CCGCGGCACTCTGCTCGCTCCCTCGGAACTACATCACCGACAGTCCGCTCTTCTTC 1092
QY 380 GlnIleValIleuArgGlnGluPheValAspSerValArgAspAspProTyR-----Gly 397
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 1093 CAGACCGGTATAGCAGCAGAGCTCATCGACGTCAGAGTCCGACATGATCATCAGGGGC 1152
QY 398 LeuPro----- 399
DB 1153 AACCTGGGATCTGGGTACAGCGTCACGTCACGCCGCGCGAGACCAACAGAC 1212
QY 400 -----TrrPrrLysGluAlaValAlaGlnHisIleAlaLysAsnProThrAspAlaLeu 417
DB 1213 CCGGACGTGTGAGAACGAAAGGTGAAGAACCATGATGACACACGAGAGACCCGCTT 1272
QY 418 ProIleProPheArgAspProGluProGlnValItrThrProPheThrGluGluHisPro 437
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 1273 CCAATCCGCTTCGAGAGACCCCGAGCGGACGACACCTTGTTCAGCCATCGACCGC 1332
QY 438 TrpHisThrGlnIleHisArgAspAlaPheSerTyRArgAlaValGlyProGluValAsp 457
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 1333 TGGCAGACTCAATTCACCGCATCGTCACTTACGGCGGCGTCAGCAAGCATCGAC 1392
QY 458 SerArgValIleValAspLeuArgTrpPheGlyAlaThrAspProGluAlaAsnLeu 477
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 1393 TCACGTTCATGTGTGATGCGGCTTCTTCGCGCGGACGACGACCAAGAGAGAAACAG 1452
QY 478 LeuValIleGlnAsnAspValGlnAspGlyTyRSerMetProGlnProThrPheArgTyR 497
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 1453 CTCTGCTCTCGGACAAATTCAGGACAGTCATCAATCCGACGCGGCTTCGACTTC 1512
QY 498 Arg---ProSerThrAlaSerAsnValArgAlaIaIaGlyIleMetMetAlaAspMetCysGlu 516
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 1513 CGCTTCGCGGCGCGGACGAGCAAGGAGGAGACATGATGACCGATGTGCGCTT 1572
QY 517 ValAlaSerAsnLeuGlyTyRLeuProThrSerProProGlnIlePheMetAspProGly 536
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 1573 ATGTGGCGAGATGTGGTCTCTGCTCCGCGCTCCCTCCCAATTCATGAGCCCGGT 1632
QY 537 LeuAlaLeuHisLeuAlaGlyThrThrArgIleGlyPheAspLysAla-----ThrThr 554
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 1633 CTGTGCTTCACCTCGGTGTGACGACGACGATGTCGACGACGAGAGAGACAAGTGC 1692
QY 555 ValAlaAspAsnAsnSerLeuValItrPAspPheAlaAsnLeuTyRValAlaGlyAsnGly 574
    ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
DB 1693 TGGCGTCAACAGCAGCTGGCGTGTGGCTTCAAGAACCTGCTTCCTGCGTGGCGGA 1752
QY 575 ThrIleArgThrGlyPheGlyLysAsnProThrIleuThrSerMetCysHisAlaIleLys 594
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 1753 AACATTCACCGGCTACGCGCGGACCAACCGACGCTACCGCAATGCTCGTCGCAACG 1812
QY 595 SerAlaArgSerIleLeuAsn 601
DB 1813 AGTTCGAGTACATCAAGAAC 1833

```

RESULT 6 D73369 2046 bp mRNA linear PLN 04-FEB-1999
LOCUS D73369
DEFINITION Corioliu versicolor mRNA for pyranose oxidase, complete cds.

ACCESSION D73369.1 GI:1845548
 KEYWORDS pyranose oxidase; PROD.
 SOURCE Coriolus versicolor
 ORGANISM Coriolus versicolor
 Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes; Apurilloporales; Coriolus.
 1 (sites)
 Nishimura, I., Okada, K. and Koyama, Y.
 Cloning and expression of pyranose oxidase cDNA from Coriolus versicolor in Escherichia coli
 J. Biotechnol. 52 (1), 11-20 (1996)
 97177816
 9025322
 2 (bases 1 to 2046)
 Nishimura, I., Okada, K. and Koyama, Y.
 Cloning, sequencing analysis, and expression in Escherichia coli of pyranose oxidase cDNA of Coriolus versicolor
 Unpublished
 3 (bases 1 to 2046)
 Nishimura, I.
 Direct Submision
 Submitted (05-OCT-1995) Ikuko Nishimura, Kikkoman Corporation, Research and Development Division: Noda 399, Noda, Chiba 278, Japan (Tel: 0471-23-5571, Fax: 0471-23-5550)
 Submitted (5-Oct-1995) to DDBJ by:
 Ikuko Nishimura
 Dept. of Research and Development Division
 Kikkoman Corporation
 Noda, Chiba 278
 Japan
 Phone: 0471-23-5571
 Fax: 0471-23-5550.
 Location/Qualifiers
 1..2046
 /organism="Coriolus versicolor"
 /mol_type="mRNA"
 /strain="ps4a"
 /db_xref="taxon:57466"
 26..1897
 /note="PROD"
 /codon_start=1
 /product="pyranose oxidase"
 /protein_id="BA11119.1"
 /db_xref="GI:1845549"
 /translation="MSTSSDPFNFNTKSPFSAAOAKASATSLPPLGPDKKVPGMD IKYVIVAGSGPICGTARELVEAGYVAMPDIGEDISGLIKGAKKNTVEYOKNIDK EVNTIQGLMVSVPNTLVITDLSPTSMQASSPFVANGSPEDDPLRNISGQAVTV VGGSTHTMTCATPFRDQRPDLVKDDQDADADAWDRLYTKAESYFTGTIDQESTIR HNVLNKLAEYKQGRDFQOQIPLAATRRSPFVEVSNANFYEDLQNPNDAPERNEN LFPVACERVANPOLLPSLGSYITEOSLVFCQVSTELDSVKDMLIRGNPDIGYS VTYRPGAGETKNHPDMNMEKYNHMMHOQEDPLIPEDDPEQVYTLFQPSHPMTQIHR RDAISYAAVQOSIDSRLLVDRFGRGRTEPEKNLMSDKITDTYNNPQTFPDRFPA GRISKEADNMNTDNCVMSAKIGFLPSLDFQEMEPGLVHLGCTHRNGFQEDKCCV NNTDSRVGEFKNLIFGGCGNIPATAGANPILAMSLAIKSEYIKNNTPSPFTDQAE"

```

Oy 398 LeuPro----- 399
Db 1178 AACCTGGCGATCTGGGGTACAGGCTCAGTACAGCGCCGGGCGGAGACCAACAGACAC 1237
Oy 400 -----TrrTrrpLysgluaIaValaIaGlnHisIleAlaLysAsnProThrAspAlaLeu 417
Db 1238 CCGGACTGTGGAGCAAAAGAGAACCATGATGAGCAGCACAGAGAGACCCGCTT 1297
Oy 418 ProIleProPheArgAspProGluProGlnValaIaThrThrProPheThrGluGlnHisPro 437
Db 1298 CCAATCCCGTTTCGAGAGCCCGAGCCGAGGTACACCTTGTCCAGCCATTCGACCCG 1357
Oy 438 TrpHisThrGlnIleHisArgAspAlaPheSerTrpGlyAlaValaIaGlyProGluValaAsp 457
Db 1358 TGGCAGACTCGATTCACCGCGGATGCGTTTCAGTTACGGCGGGGTGCAGCAAGCATGCAC 1417
Oy 458 SerArgValaIleValaAspLeuArgTrpPheGlyAlaThrAspProGluAlaAsnAsnLeu 477
Db 1418 TCAGTCTCATCTGCTGAGCTGCTTCGCGCCGAGCAGCAAAAGAGAGAGAACAG 1477
Oy 478 LeuValaIaPheGlnAsnAspValaGlnAspGlyTrpSerMetProGlnProThrPheArgTrp 497
Db 1478 CTCTGGTCTTCGCAAAATTACGACACGTACACATGACCGCAGCCGAGCTTCGACTTC 1537
Oy 498 Arg---ProSerThrAlaSerAsnValaArgAlaArgLysMetMetAlaAspMetCysGlu 516
Db 1538 CGCTTCCGGGGGCGGCGAGCAGCAGAGGCGGAGCATGATGACCGATATGCGCTT 1597
Oy 517 ValaIaSerAsnLeuGlyGlyTrpLeuProThrSerProPropGlnPheMetAspProGly 536
Db 1598 ATGTCGGGAGAAATGTTGGCTTCCTCCGCGCTCCCTCCCAATTCATGAGAGCCGCT 1657
Oy 537 LeuAlaLeuHisIleAlaGlyThrThrArgIleGlyPheAspLysAla-----ThrThr 554
Db 1658 CTTCCTCTTCACCTCGGTGGTACGACACCGCATGGCGCTTCGACGAGCAGAGACACAAGTGC 1717
Oy 555 ValaIaAspAsnAsnSerLeuValaIaTrpAspPheAlaAsnLeuTrpValaIaGlyAsnGly 574
Db 1718 TGCCTCAACAGCAGACTCGCGCGGTGGTTGGCTTCAAGAACCGTTCCTCGGTGGCGCGGA 1777
Oy 575 ThrIleArgThrGlyPheGlyGlnAsnProThrLeuThrSerMetCysHisAlaIleLys 594
Db 1778 AACATCTCCACCGGCTGCGGCGGACCGGAGCTCACCGCAATGTCGTCGCGATCAAG 1837
Oy 595 SerAlaArgSerIleIleAsn 601
Db 1838 AGTTCGAGTACATCAAGAAC 1858

RESULT 7
E32737 1902 bp DNA linear PAT 18-JUN-2001
LOCUS Panus cornucopiae-derived antitumor protein and gene thereof.
DEFINITION E32737
ACCESSION E32737.1 GI:13017467
VERSION JP 1999315096-A/1.
KEYWORDS Pleurotus cornucopiae (cornucopia mushroom)
SOURCE Pleurotus cornucopiae
ORGANISM Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
Agaricales; Pleurotaceae; Pleurotus.
REFERENCE 1 (bases 1 to 1902)
AUTHORS Akhito,M., Koji,I., Kazuaki,K., Tomohide,S. and Ryoko,A.
TITLE Panus cornucopiae-derived antitumor protein and gene thereof
JOURNAL Patient: JP 1999315096-A 1 16-NOV-1999;
NYUFUDO KURIESHON GIUTTSUKENKUKUMITAI
OS Pleurotus cornucopiae
PN JP 1999315096-A/1
PD 16-NOV-1999
PF 07-AUG-1998 JP 1998236349
PR
PI AKHIRO MORITA, KOJI IZUMO, KAZUAKI KIDO, TOMOHIDE SAKA, PI
RKO KO AO KI
PC C07K14/375, A61K35/84, A61K38/00, C07K16/14

```

```

CC FH Key Location/Qualifiers
FT source 1..1902
FT /organism='Pleurotus cornucopiae'.

FEATURES
    source
        location/Qualifiers
            1..1902
            /organism='Pleurotus cornucopiae'
            /mol_type='genomic DNA'
            /db_xref='taxon:5321'

BASE COUNT 437 a 556 c 479 g 430 t
ORIGIN

Alignment Scores:
Pred. No.: 7.99e-87 Length: 1902
Score: 1232.00 Matches: 275
Percent Similarity: 57.05% Conservative: 77
Best Local Similarity: 44.57% Mismatches: 193
Query Match: 37.52% Indels: 72
DB: 6 Gaps: 18

US-09-856-327-2 (1-618) x E32737 (1-1902)
Oy 18 IleAsnGlyGlnIleProLysAsnAlaIleHisIleGluThrTrpGlyAsnAspGlyValaAsp 37
Db 130 ATTCGGGGAACCTTCGTCGGA-----GATACATCCAGTTCGAC 171
Oy 38 ValPheIleAlaGlySerGlyProIleGlyAlaThrTrpAlaLysLeuCysValaGlna 57
Db 172 GTTGTCATCTGCGGCTCGGCCAGTTGGCTCTACTTATGAGCCGCTCTCGTCGAGGCA 231
Oy 58 GlyLeuArgValaIleMetValaGluIleGlyAlaAlaAspSerPheTrpAlaValaAsnAla 77
Db 232 GGTTCACAAAGTAGATGATTTGAATTCGAGAAATCGACTCT----- 273
Oy 78 GluGluGlyThrAlaValaProTrpValaProGlyTrpHisLysLysAsnGluIleGluPhe 97
Db 274 -----GGCCCGCAGACTT-----GTTCCCATTAAGAAAAATACGTTGAGTAC 315
Oy 98 GlnLysAspIleAspArgPheValaAsnValaIleLysGlyAlaLeuGlnIleValaSerVal 117
Db 316 CAGAGACATTTGACAAATTCGTCATCTTCAAGGACAAATCATCTGCTGATCCGTC 375
Oy 118 ProValArgAsnGlnAsnValaProThrLeuAspProGlyAlaTrpSerAlaProProGly 137
Db 376 CCTGTCAACAATATGTTGCCGATACCTCTAGCCCGCTACTGCGAGGCC-----TCA 429
Oy 138 SerSerAlaIleSerAsnGlyLysAsnProHisIleGlnArgGluPheGlnAsnLeuSerAla 157
Db 430 ACACACTTGTCCGACGCGCAAAACCCGAGCAAGATCCCTTACTAATCTCTCAAGC 489
Oy 158 GluAlaValaIleThrArgGlyValaGlyLysMetSerThrHisIleProThrCysSerThrPro 177
Db 490 CAAGCTGTGACCCGCTGTGGAGAAATGGCGAGCACTGACTGTGCGAACACTGCG 549
Oy 178 IleHisProPheMetGluSerLeuProGlyIleGlyArgProLysLeu-----SerAsnAsp 196
Db 550 TTCACAAGAGT-----GAGCGCGCTTAAGCTTGAAGATAT 588
Oy 197 ProAlaGluAspAspLysGluTrpAsnGluLeuTrpSerGluAlaGluArgLeuIleGly 216
Db 589 GACTCCGCGGAGCGCATGAGTGGAGACGCTCTATGACATCGCCGAATCTTGCTGTCAMG 648
Oy 217 ThrSerThrLysGluPheAspGluSerIleArgHisThrLeuValaIleuArgSerLeuGln 236
Db 649 ACCGGGCTAATCTAGTTGATCAATCTATCCGGCAACACTTGTACTGAGAACCTTAG 708
Oy 237 AspAlaTrpLysAspArgGlnArgIlePheArgProLeuProLeuAlaCysHisArgLeu 256
Db 709 GAGTCTCATGCAAGT---CAGCGGGAATTTAGAGAGATCCCTCTGGCGGCCAACGCGAC 765
Oy 257 LysAsnAlaProGluTrpValaGluTrpHisSerAlaGluAsnLeuPhe-----HisSer 274
Db 766 AAC-----CCCGCTTCGTCGAGTGGAGCTGCGCATAGGTCCTTGAACCTTGAGAAC 819

```

OY	275	ILEYRASNAPRSPGLNLYSLYSLEUHEPHTIRLEUHTIRASNHIASRGCSYTHR	294
DB	820	CGCCCAACGGCGGATGACGACAGAGGGCGTTTCAACTCTTCCCGCAGTTGTATGTGAA	879
OY	295	ARGLEUALALEUTHRGILYTRGLULYSYSLIEGLIYALALAGLUALARGASNU	314
DB	880	CGCGTTACG---CGGATTTCCCTGCGACCGGAATAGAGCAATTTAGGTCATCACTCG	936
OY	315	LEUALHTIRAGSAMPROSERSECDINLEUASPSETRYR---ILEMEALILYVALTYR	333
DB	937	ATTAGCGC-----GATCGCTCAAGGTCGTAAGCGCTGACGCTGTT	975
OY	334	VALLEUALASERGLIYALILEGLIYASNPROGLINILEUTRYASNSERGILYPHESERGLY	353
DB	976	ATCCTTTGTCTGGCGAGCTCCATATATCCCAATTTCTTTGAACCTGTGATTC---GGG	1032
OY	354	LEUGINVALTHIRPROIAGSNAPSP-----SERLEULIEPROASNLEUGILYARG	369
DB	1033	CGCATGGGTGCACGAGATTCTTCACTGCCACGCCGCACTCGCTGCTCATCTGTGTGAGC	1092
OY	370	TYRILIEPHTIRGLUGLNPROMETALAPHECYSGINILEVALIEUATRGILGLUPHIVALSP	389
DB	1093	TACATCACTGAGCAGACACTCAGCTGTTCGCCAGACCGTTCAGCACCAGAACTGTCAAT	1152
OY	390	SERVALIARGASAPRPRO---TYRGLYLEUPRO-----	399
DB	1153	CTTGTCAAGTGGACATGATTATTGTGTGGACGCCGCCGACAGCAGACTACAGTGTACT	1212
OY	400	-----TRTPRILYSGIUALIYALAGLIN	407
DB	1213	TTACACCCCGGACAGCTCGAGCAATPACAGCCCAATTTGTGTGAACAGAGCTCAGAG	1272
OY	408	HISILIEALYASNPPIROTHIRASPALALEUPROILEPHTOPHEATARGSPROGLIAPROGLIN	427
DB	1273	CACATGATGACGACACACGAGAGATCCCTCCGATCCGCGTCGATGATACCTGAGCCACAG	1332
OY	428	VALTHIRPIROPHETHTIRGLUHSIPROTPIRPHSTHRTGILNILEHSIARGSPALAPHE	447
DB	1333	GTCACGACACTCTTCCAGGACACACTCAGTGCACACTCAGATTCACTCGTGAATGCTCTT	1392
OY	448	SERTYRGLIYALVALGLYPROGLIVALIASERIRGVALILEVALSPLEUATIRPHE	467
DB	1393	AGCTACGGCGCTGTGGCGGAGATATGACAGCGCTGTGTGTGATGACTGGCGTCTCTC	1452
OY	468	GLYALATHIRAPPROGLIUALIASNAPSNLEUVALPHEGLINASNAPSPVALGLNASPLY	487
DB	1453	GGTGGCACTGGAACCTGTGGAGAGAACGACTGTGTGTTCTCTCAAAACAGATCACTGATGCA	1512
OY	488	TYRSEMETPROGLINPIROTHIRPHEARTGYRIRG---PROSEHTIRALIASERASNVALARG	506
DB	1513	TACAACTGCCACAGCCCACTTACAGCTTCGCGTTCCCCCAGGGCCGACAGCCACAGAG	1572
OY	507	ALAAARGLYSMETMETALIASPMETCYSGIUALIASERASNLEUARGTYRLEUPRO	526
DB	1573	GCCGAGCTCATGATGGCTGACATGTCGACAGATGTCACAAAGTCGGGCTTCCTGCCC	1632
OY	527	THIRSEPIROPOGLINPHEMETASPPROGLIYLEUALALEUHSILEUALIAGLYTHIRTHARG	546
DB	1633	GGTTCATATCCGCAAGTTTATGGCTCTCGGCTGTGTACTGCCACTGTGGTGGAAACCCAGC	1692
OY	547	ILEGLYPHEASRPLYSALATHIRTHIVALIA-----ASPASNASERLEUVALITRASP	564
DB	1693	ATGGGCTTGTATGAGAGAGCTGATAGAGCGTGTGTGCACACCAACCTCAAGGTCTTTCGCC	1752
OY	565	PHEALASNPLEUTRYVALIAGLYASNGLIYTHRILEARTGHRGLYRPHIEGLIYLUASNPRO	584
DB	1753	ATGGAGAACCTGTTCTGTGGGTGGCGGCAACATTTGGCACCGGGTATGCTCGAACCCG	1812
OY	585	THIRLEUPHTIRSEMETCYSHSIALIYLEUSERLALIASERILELLEASN	601
DB	1813	ACGCTACTGGGTGGCTGGCTTGCATATGAGAGCTGTATGATATACAGGAT	1863

LOCUS	AR141573	1869 bp	DNA	linear	PAT 08-AUG-2001
DEFINITION	Sequence 1 from patent US 6146865.				
ACCESSION	AR141573				
VERSION	AR141573.1				
KEYWORDS	GI:15101089				
ORGANISM	Unknown.				
REFERENCE	Unknown.				
AUTHORS	1 (bases 1 to 1869)				
TITLE	Christensen, S., Lassen, S., Flensted, and Schneider, P.				
JOURNAL	Nucleic acids encoding polypeptides having pyranose oxidase activity				
FEATURES	Patent: US 6146865-A 1 14-NOV-2000;				
SOURCE	Location/Qualifiers 1. 1869				
BASE COUNT	397 a 622 c 548 g 302 t				
ORIGIN	/organism="unknown"				
Alignment Scores:					
Score. No.:	4.09e-85	Length:	1869		
Percent Similarity:	1210.00	Matches:	274		
Best Local Similarity:	56.19%	Conservative:	71		
Query Match:	44.63%	Mismatches:	197		
DB:	36.85%	Indels:	72		
	6	Gaps:	16		
US-09-856-327-2 (1-618) x AR141573 (1-1869)					
QY	37 AspyAlPheIIeaIIaGlySerGIyProIIleGIyAlaThrTyraLaLysLeuCyVaIGlu	56			
Db	142 GACGTTGCATCGTCGCGCTGCGGCCGCGATGTGCGACATATATGCGCGCGAGCTGTGAG	201			
QY	57 AlAGlyLeuArGyValMetValGIuIIleGIyAlaLaAspSerPheTyraLaValAsn	76			
Db	202 GCCGCTTCACAGTCGCGCATCTTCGAGATTGGAAGATCGACCTCCGGCTTGAGATC--	258			
QY	77 AlaGluGluGIyThraLaValProTyraValProGIyThraLysLysAsnGIuIIleGIu	96			
Db	259 -----GCGTCACACAGAAAGAACCCGCTGAG	285			
QY	97 PheGluAspPheAspPheValAsnValIIleLysGIyAlaLeuGluGluValSer	116			
Db	286 TACACAGAAACATCGACAAATTCGTAATGTATACAAAGGCAACTATGCCGCTCG	345			
QY	117 ValProValArGAsnGluAsnValProThrLeuAspProGIyAlaTyrPserAlaProPro	136			
Db	346 GTGCCCGTCACACAGATGTCGTGTGACACGTTAAGCCGGCGTCATGGCAAGCT-----	399			
QY	137 GlySerSerAlaIIleSerAsnGIyLysAsnProHISGluGluGluPheGluAsnLeuSer	156			
Db	400 TCGACGTTCTTCGTCGCCGACGGGGCGCATTCACAGCAAGACCCGTCGCCCACTTAGT	459			
QY	157 AlaGluAlaValThrArgGIyValGIyAlaLysSerThrHISThrPheValIIleAsnLeuSer	176			
Db	460 GCCACGCGCGTCACCCCGCGTCGTGGCGCGCATGTCCTACGCATGAGCGTGGCGACCCG	519			
QY	177 ArgIIleHISProPrometGluSerLeuProGIyIIleGIyArGProLysLeu--SerAsn	195			
Db	520 CGCTTCGAGAG-----CTGACAGCCCGCGTCGTCGTGAAAGAC	558			
QY	196 AspProAlaGluAspAspLysGIuTTPAsnGluLeuTyrserGIuAlaGluGluLeuIIle	215			
Db	559 GACATCCAGCGCGGAGACGCGCATGTGGACGGCTTCACAGAGGCCGCAATCGTACTTC	618			
QY	216 GlyThrSerThrLysGluPheAspGluSerIIleArgHISThrLeuValIIleAsnLeuSer	235			
Db	619 AAGACGCGGACGACCCCACTTGGCCGAGTCGATCCGACCAACACTCGTCTGCTAACAACCTG	678			
QY	236 GluAspAlaTyrsAspArgGluArgIIlePheArProLeuPheProLeuAlaCysHISArg	255			

Db 679 CAGGAGAGTACAAGGC---GTGCGGCACTTCCAGAGATCCGCTCGCGGCGAGCGGC 735
 Oy 256 LeuysasnaIaProgluYrValIgluThrhIsSerIaIaIuAsnLeuPhe-----His 273
 Db 736 -----CAGAGCCCGAGCTTGTGAGTGGAGCTCGGCGCACACCGCTGTGATCTCGAG 789
 Oy 274 SerIleYrAsnAspRlySgInIlySLeuPheThrhLeuThrhAsnIaArgCys 293
 Db 790 AACCGGCGGAGAACAGAGCGCGGAGAGCGCTTCAACSTCTTCCCGCGCTCGCGTGC 849
 Oy 294 ThrArgLeuAlaLeuThrhIleYlYrGluYsIleGlyAlaIaIaGluValArgAsn 313
 Db 850 ACGAGCTGAGCGCGGATACGCGACSTCGAGATCTGAGC---CTGGAGTCTCGCGAC 906
 Oy 314 LeuLeuAlaThraArgAsnProSerSerGlnLeuAspSerYrIleMeAlaIaValYr 333
 Db 907 CTCACGCGGGGCAAGACATCAC---ATCAAGCGCAAGGTTGAC 948
 Oy 334 ValLeuAlaSerGlyAlaIleGlyAsnProGlnIleuYrAsnSerGlyPheSerGly 353
 Db 949 ATCTCAACCGCGCGGCTCCACACGCGCAGCTCTCGCGCTGTGGATTGGGCGAG 1008
 Oy 354 LeuGlnValThrProIArgAsnP-----SerIlePProAsnLeu 367
 Db 1009 CTG-----GGTCGTCCCGACCGCGCAAGCGCTGCCGTCTGTCTCGCTACCTG 1059
 Oy 368 GLYArgYrIleThrhIgluInProMeAlaPheCysGlnIleValLeuArgGlnGluPhe 387
 Db 1060 GGGACCCACATCGACCGAGACAGCTGTCTGTCCACAGCTCATGAGCAGCAGAGCTC 1119
 Oy 388 ValAspSerValaArgAspPro-----TyrlYleuPro----- 399
 Db 1120 ATCAACAGTGTACACCGGAGATATGACCATTTCTCGCAAGCCCGCGACCTATAGC 1179
 Oy 400 -----TrrYrIySgIuaIaVal 405
 Db 1180 GTCACTATATCCCGGCGCAACCCGACACACAGCAGCCGAGCTGTGGAGAGAGAGTG 1239
 Oy 406 AlGlnHisIleAlaIySAsnProThraSpAlaLeuProIleProPheArgAspProGlu 425
 Db 1240 AAGAGACATGATGACCCACAGAGAGACCCGCTCCGATCCGCTCGAGAGACCTGAG 1299
 Oy 426 ProGlnValThrThrProPheThrhIgluIuhIsProThrhIsthGlnIleIaArgAsp 445
 Db 1300 CCGAGGTCAACAGCGCTTTCAGAGCAGCAGCACCAGTGGCAGACCATTCACCGCGAC 1359
 Oy 446 AlaPheSerYrGlyAlaValGlyProGluValaSpSerArgValIleValaSpLeuArg 465
 Db 1360 GCCTTCACTACGGCGCGCTGCACAGACATCGACTCGCGCTCATCTGCTGACTGGCGG 1419
 Oy 466 TrpPheGlyAlaThraSpProGluAlaAsnAsnLeuValaPheGlnAsnAspValGln 485
 Db 1420 TTCTTCGAGCGACCGAGCCAGAGAGAGACAGATATGCTTCTGGAGACATACG 1479
 Oy 486 AspIlyYrSerMetProGlnProThrhPheArgYrArgProSerThrhAlaSerAsnVal 505
 Db 1480 GAGCGGTACAACCTCCGAGACCGACGTCGACTTCGCG-----TTCCCGCGGGGGCGCG 1533
 Oy 506 ArgAlaArgIySmetMetAlaAspMetCysGluValaIaSerAsnLeuGlyIlyrLeu 525
 Db 1534 GAAGCGGAGAGACATGATGACCGATGTGCTCATGTGCGGAGAGATCGGTGGATTCTCG 1593
 Oy 526 ProThraSerProProGlnPheMetAspProGlyLeuAlaLeuHIsleuAlaGlyThrThr 545
 Db 1594 CTGGGTCTTACCAACAGTTCATGAGCCGCTTCTGTCTGCACTGTGTGGAGACGAC 1653
 Oy 546 ArgIleGlyPheAspIlySalaIaThr-----ThraValaIaAspAsnSerIleYr 563
 Db 1654 CGCATGGCTTCGAGAGAGCGGACAGTGTGCTGCGTGCAGACAGCACTCAGCGTCTTC 1713
 Oy 564 AspPheAlaAsnLeuYrValaIaGlyAsnGlyThrhIleArgThrhIyPheGlyGluAsn 583
 Db 1714 GCGTTCAAGAACCTTCTCGCGGCTGGGGAACATCCACCGGTTACGCGCGGAC 1773

Oy 584 ProThraThrhSerMetCysHisAlaIleIySserAlaIaArgSerIleIleAsnThrhLeu 603
 Db 1774 CCGAGCGTCAACCGGATATGCTTGTGATTCAGAGACTCTCGAGTACATC-----AAG 1824
 Oy 604 LysGlyGlyThraSpGlyIySAsnThrhIgluIuhIsArgAsn 617
 Db 1825 AAGAACTTCGAGCGGAGCGGACCGGACCCGCTGAGAGCAGCAGAAC 1866
 RESULT 9
 ARI70485
 LOCUS ARI70485 1701 bp DNA linear PAT 17-DEC-2001
 DEFINITION Sequence 2 from patent US 6291648.
 ACCESSION ARI70485
 VERSION ARI70485.1 GI:17908444
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 1701)
 AUTHORS Kawamura,Y., Morita,A., Izumo,K. and Saka,T.
 TITLE Antitumor protein and corresponding gene sequence isolated from
 matsutake mushrooms
 JOURNAL Patent: US 6291648-A 2 18-SEP-2001;
 FEATURES Location/Qualifiers
 source 1..1701
 BASE COUNT 411 a 463 c 432 g 395 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 7,24e-76 Length: 1701
 Score: 1091.00 Matches: 248
 Percent Similarity: 57.04% Conservative: 84
 Best Local Similarity: 42.61% Mismatches: 184
 Query Match: 33.22% Indels: 66
 DB: Gaps: 15
 US-09-856-327-2 (1-618) x ARI70485 (1-1701)
 Oy 37 AspValPheIleAlaIySerglyProIleGlyAlaThrYrAlaIySLeuCysValGlu 56
 Db 97 GATGTTTCATTTGCTGGCAGTGTCCATAGCTAGTACTACTAGCCCGCCACATCATTTGAC 156
 Oy 57 -----AlaGlyLeuArgValaValaMetValGluIleGlyAlaIaIaAspSerPheYrAla 74
 Db 157 AATACCTCAACTACAAGGTTTACATGGCCAAATAGGTTTCAAGATAC----- 207
 Oy 75 ValAsnAlaIgluIgluYlThraIaValProYrValProGlyYrHIsIySAsnGlu 94
 Db 208 -----CTGTGATC---GGGGCCCATCACAGAACCTCC 237
 Oy 95 IleGluPheGlnIySAspIleAspArgPheValaIySValIleIySgIyAlaLeuGlnGlu 114
 Db 238 ATAAAGTTTCAGAAAGACATTCAGACAGTTTGTGAATATCATCAACGCGTCCACGCG 297
 Oy 115 ValSerValProValaIySAsnGlnIySValaIyProThrhLeuAspProGlyAlaIaTrpSerAla 134
 Db 298 ATTTCGATTTCGCCATGCGACACTTACCAGCCCACTCGCTGAGCAGCGTGG---GCG 354
 Oy 135 ProProGlySerSerAla-----IleSerAsnGlyIySAsnProHIsGlnArg 150
 Db 355 CCGGCCATCGATCTCTCCGGAAGCGCAGCTCTGATTATGGACACAAATCCGAATCAGAG 414
 Oy 151 GluPheGluAsnLeuSerAlaIgluAlaValaIyThraArgIyValaIyGlyIyMetSerThrhIs 170
 Db 415 GCCGCGCTCAACTTCCCGGTAGCGCTGTCACTAGAGACAGTGGGGGATGCGCAGCCAC 474
 Oy 171 TrpThrcysSerThrhProArgIyIleHIsProIleMetGluSerLeuProGlyIleGlyArg 190
 Db 475 TGGACTTGGCGGTCTCTACTCCACATGAC----- 504
 Oy 191 ProlYsLeuSerAsnAspProAlaIgluAspAspIySgluTrpAsnGluLeuYrSerGlu 210

```

Db      505 GAGAGAGGGGTCAACACCCAGTT---GACACAGCAGGAGTTGCAGCCATGCTCGAACGT 561
Oy      211 AlAGluArgLeuIleGlyThrSerThryLysGluPheAspGluSerIleIleArgHisThrLeu 230
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      562 GCTAAACACTTGTCTCAGCTTCCACAGCCAGCAGTACAGCAGATTCTATCCGTAGATAGTT 621
Oy      231 ValLeuArgSerLeuGlnAspAlaTyrLysAspArgGlnArgIlePheArgProLeuPro 250
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      622 GTCAAGAGACTCTTCAGCAGACCCTT---GATCGCTCGGGGGTGTGACCACTCCCGG 678
Oy      251 LeuAlaCysHisArgLeuLysAsnAlaProGluTyrValGluThrPheSerIleGluAsn 270
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      679 CTGGGGGGTGGCCCGTACGACCAATCCTATTATGTACCTGACCGGCTCCGAT--- 735
Oy      271 LeuPheHisSerIleTyrAsnAspAspLysGlnLysLysLeuPheThrLeuThrAsn 290
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      736 -----ACCGTCTTGCTGATGTGCCGAGAGAGTCCCGGATTCGCTTGGTTACAGAG 786
Oy      291 HisArgCysThrArgLeuAlaLeuThrGlyTyrGluLysLysIleGlyAlaAlaGlu 310
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      787 ACGAGAGTGCAGACCTTATGTGTCAAGTAACCAATCCGACGAGGTGTCTGCTGCTG 846
Oy      311 ValArgAsnLeuLeuAlaThrArgAsnProSerSerGlnLeuAspSerTyrIleMetAla 330
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      847 CTACGTAACCTG-----AATACACAGCAAC-----GATGAACTTGTCTGGGCC 888
Oy      331 LysValTyrValLeuAlaSerGlyAlaIleGlyAsnProGlnIleLeuTyrAsnSergly 350
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      889 AAGACTTTCGTCATAGCTGTGTGACAGAGTCGACACCGCAATCTGTGGAAACG---- 945
Oy      351 PheSerGlyLeuGlnValThrProArgAsnAspSerLeuIleProAsnLeuGlyArgTyr 370
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      946 -----ACATCCGCCCATAT-----GCCGTTGTGTGCTGCT 975
Oy      371 IleThrGlnGluProMetAlaPheCysGlnIleValLeuArgGlnGluPheValAspSer 390
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      976 CTCACGAGACAGTCCATGACTTTTGTGATGATGTTCTCAAGAGGGGCACTAGTCAGTGC 1035
Oy      391 ValArgAspProTyrGlyLeuProTyrTrpLysGluAlaValAlaGlnHisIleAla 410
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1036 ATCGCTAGTACCCCTGCG-----TTCGCTGGCAGAGTTGAGCGCGCAACAAG 1083
Oy      411 LysAsnProThrAspAlaLeuProIleProPheArgAspProGluProGlnValThrThr 430
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1084 AAGCACCCGATGAGTGTGCTCCATTCATCCAGACCCGAACTTCAGTCAATGATGAT 1143
Oy      431 ProPheThrGlnGluHisProThrPheThrGlnIleHisArgAspAlaPheSerTyrGly 450
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1144 CCGTACAGCTCGGACTTCCCTTGGCATGTTCAGTGATGCGGATGCAATTCATATGCT 1203
Oy      451 AlaValGlyProGluValAspSerArgValIleValAspLeuArgTrpPheGlyAlaThr 470
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1204 GATGTTGGACCCAGCGGACCCGGGTGTGTGTCGATGCTGAGGTTTTCGCAATCA 1263
Oy      471 AspProGluAlaAsnAsnLeuValPhe-----Gln 481
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1264 GATATTGTTCGAAGAAATCGAGTACTTCCGCTCCGAACCTTAAGCTACGCGAGTGGGA 1323
Oy      482 AsnAspValGlnAspGlyTyrSerMetProGlnProThrPheArgTyrArgProSerThr 501
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1324 GCGGGTGTTCACACACTTATGGAATGCCACAGCCGACATTCATGTCAGGAGGCCAAC 1383
Oy      502 AlaSerAsnValArgAlaArgLysMetMetAlaAspMetCysGluValAlaSerAsnLeu 521
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1384 GCCGATGAGACCGGTCGACAGAGATGATGATATGATACCAAGCTCCGCAACATGCTG 1443
Oy      522 GlyGlyTyrLeuProThrSerProGlnPheMetAspProGlyLeuAlaLeuHisLeu 541
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1444 GGTGGGTACCTTCGTGGCTCAACCTCAATTTATGAGCAGCTGCTGCTACTGCAACATC 1503
Oy      542 AlaGlyThrThrArgIleGlyPheAspLysAlaThrThrValAlaAspAsnAsnSerLeu 561
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

Db      1504 ACGGAACTACTCCGATCGGAGACAGATGCAAACTTCTGTTCTGATCCGACATCMAAG 1563
Oy      562 ValTrpAspPheAlaAsnLeuTyrValAlaGlyAsnGlyThrIleArgThrGlyPheGly 581
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1564 GTTCATTAATCTCAACATCTGTGGGTGGCGGGGAGTGGTGGATTCACATCGCACTGCC 1623
Oy      582 GluAsnProThrLeuThrSerMetCysHisAlaIleLysSerAlaArgSerIleLeuAsn 601
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1624 TCCAAACCCGACTCGTACGAGCGCTGCGGTATGCGCTCAAGGGTGTGAGCTGTACTCAT 1683
Oy      602 ThrLeu 603
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1684 TACCTT 1689

RESULT 10
E33665
LOCUS      1946 bp      DNA      linear      PAT 31-JAN-2002
DEFINITION Novel protein, gene encoding the same and method for utilization
            thereof.
ACCESSION  E33665
VERSION    E33665.1 GI:18624149
KEYWORDS   JP 2000083675-A/1.
SOURCE     Tricholoma matsutake
            Tricholoma matsutake
            Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
            Agaricales; Tricholomataceae; Tricholoma.
REFERENCE  1 (bases 1 to 1946)
            Takakura,Y., Kuwata,S. and Ota,S.
            Novel protein, gene encoding the same and method for utilization
            Patent: JP 2000083675-A 1 28-MAR-2000;
            JAPAN TOBACCO INC. SOCIETY FOR TECHNO-INNOVATION OF AGRICULTURE
            FORESTRY AND FISHERIES
COMMENT    OS Tricholoma matsutake
            PN JP 2000083675-A/1
            PD 28-MAR-2000
            PE 08-SEP-1998 JP 1998270606
            PR YOSHIMITSU TAKAKURA, SHIGERU KUWATA, SYOZO OTA
            PT C12N15/09,A01N37/12,A01N63/00,A01N63/02,A01N65/00,C07K1/14,PC
            C07K1/18,
            PC C07K1/30,C07K14/375,C12N1/19,C12N1/21,C12N5/10,C12N5/10,PC
            C12P21/02,
            PC C12Q1/68//((C12N15/09,C12R1:645),C12N15/00,C12N5/00,C12N5/00,
            PC (C12N15/00,C12R1:645)
            CC CC
            FH Key Location/Qualifiers
            FT source 1..1946
            FT 1..1946 Location/Qualifiers
            FEATURES
            source 1..1946 Location/Qualifiers
            /organism="Tricholoma matsutake"
            /mol_type="genomic DNA"
            /db_xref="taxon:40145"
BASE COUNT 484 a 516 c 475 g 471 t
ORIGIN
Alignment Scores:
Pred. No.: 4,13e-72 Length: 1946
Score: 1044.00 Matches: 241
Percent Similarity: 55.91% Conservative: 90
Best Local Similarity: 40.71% Mismatches: 187
Query Match: 31.79% Indels: 74
DB: 6 Gaps: 17

US-09-856-327-2 (1-618) x E33665 (1-1946)
Oy      27 IleHisGluThrTyrGlyAsnAspGlyValAspValPheIleAlaGlySerGlyProIle 46
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      185 GTACATTCACACT-----GATGTTTCATGTGCGGAGTGTCCCAT 226
Oy      47 GlyAlaThrTyrAlaLysLeuGlyValGlu-----AlaGlyLeuArgValAlaMetAl 64
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      227 GCTGTACTTACGCCCGCACATCAATGATGCAATACCTCAACAAAGGTCTACATGCC 286

```


[illegible]

OY	421	PhenylarginylproglinalaValThrProthetrigluHisProthriHistr	440
Db	1214	TTCCACGACCGTGAACSTCAAGTATGATGTCCTCCGACAGTGGACSTCCCTGGCAGTGT	1273
OY	441	GlnIleHisArgAspAlaPheSerTyrGluValAlaGlyProGluValAspSerArgVal	460
Db	1274	CAGGTCCATCGSC-----TATGCAATTGGTGATGTTGTGAGCCCAAGGCCGACCGCGTGT	1327
OY	461	IleValAspIleuArgTrrPheGlyAlaThrAspProGluAlaAsnAspIleuValPhe	480
Db	1328	GTCTCGATGTCGAGGTTTTCGGCAATCGAGATTGTGTCCAGAAATCGAGACTTTC	1387
OY	481	-----GlnAsnAspAlaGlnAspGlyTyrSerMetPro	491
Db	1388	GGTCCGAAACCTTAAGSTACGCGACTGGGAAACGGGTGTTCAGACACTTATGAGATGCCA	1447
OY	492	GlnProThrPheArgTyrArgProSerThrAlaSerAsnValAlaGalaArgIysMetMet	511
Db	1448	CAGCGACATTCATCCATGTCAGCGGACCCAGCCGATGGAGACCGTGAACAGAGATGATG	1507
OY	512	AlaAspMetCysGluValAlaSerAsnIleuGlyGlyTyrIleuProThrSerProGln	531
Db	1508	AATGATATGACCAACGTGCGGAAACATCATGTGGCGGGTACCTTCCTGGCTCTACCCCTCAA	1567
OY	532	PheMetAspProGlyIleuAlaIleuHisIleuAlaGlyThrArgGlyIleuPheAspIrys	551
Db	1568	TTTATGGACACTGGTCTGCGACACAGCACTCCGGAACTACTCGATCGGAGCAAGATCAT	1627
OY	552	AlaThrThrValAlaAspAsnAspSerIleuValTrrAspPheAlaAsnIleuTyrValAla	571
Db	1628	CAAACTCTGTGTGTGATCCGACACACTCAAAAGTTTATAACTTCGACAAATCTTGGGTGGCC	1687
OY	572	GlyAsnGlyThrIleArgThrGlyPheGlyGlyAsnProThrIleuThrSerMetCysHis	591
Db	1688	GGGAGATGGGTGCATTCCAGATGCACTGCTCCGCAACCCGACTCGTACGAGCGTGGCAT	1747
OY	592	AlaIleuIysSerAlaArgSerIleIleAsnProIleu	603
Db	1748	GCGCTTAAGGCTGACGCTGTAGTACATTAACCTT	1783

FEATURES

```
source      1..1946
            /organism="Tricholoma matsutake"
            /mol_type="genomic DNA"
            /db_xref="taxon:40145"
```


Db 1448 CACCCGACATTCATGCAAGCGGACCAACGCCGATGAGACCCGTACACCGAGAGATGATG 1507
 QY 512 AlAaspwctcsgluValAlaSerAsnIeuGlyIyTrIeubProThrSerProPcglN 531
 Db 1508 AATGATATGACCAACGCGGACATCATCTGGCGGGTACTCTCTGCTCTCTACCTCTCA 1567
 QY 532 PhemeArasProclIyLeuAlaLeuHisLeuAlaGlyThrTrArqIlegIyPheAsPlys 551
 Db 1568 TTTATGGACACTGTCTCGACGACATCAGCGGAACATCTCGGATCGGACGACATGAT 1627
 QY 552 AlAtThrValAlaAspAsnSerIeuValITrpAspPheAlaLeuTyTrValAla 571
 Db 1628 CAACACTTCTGTCTGATCCGACATCAAGTTCAATCTTCACATCTCTGTGGTCCGC 1687
 QY 572 GlyAsnGlyThrIleArqThrGlyPheGlyGluAsnProThrIleuThrSerMetCysHis 591
 Db 1688 GGGAAATGGTGCATCGATGCGACTGCTCGACACCGACCTGTACGACGTCGCCTAT 1747
 QY 592 AlAlIleGlySerAlaArqSerIleIleAsnThrIeu 603
 Db 1748 GCGCTTAAGGCTGCTGAGCTGTATGTCATGTTACCTT 1783

RESULT 13
 ANIS07836 165137 bp DNA linear BCT 12-MAR-2003
 LOCUS ANIS07836
 DEFINITION Arthrobacter nicotinovorans PA01 plasmid sequence.
 ACCESSION AJ507836
 VERSION AJ507836.1 GI:25169022
 KEYWORDS 2,6-dihydroxypyridine hydroxylase; 2-keto-gluconate dehydrogenase-like protein; 6-HNO gene; 6-HNO gene; 6-hydroxy-D-nicotine oxidase; 6-hydroxy-L-nicotine oxidase; amine oxidase; amino acid permease; amino acid transporter; ATP binding component of ABC transporter; ATP-binding protein; ATP/GTP-binding protein; beta-glucosidase; carbon monoxide dehydrogenase subunit D; carbon monoxide dehydrogenase subunit E; carbon monoxide dehydrogenase subunit F; carbon monoxide dehydrogenase subunit G; CGL gene; chromosome partitioning protein; conjugative transfer gene complex protein; cox2 gene; cox3 gene; cox6 gene; cox6 gene; CSCA gene; DarnH3 gene; dhph gene; DNA helicase-like protein; DNA ligase; DNA recombinase; ebrA gene; ebrB gene; endopeptidase; ESTa gene; formyltetrahydrofolate deformylase; glycine cleavage system T protein; hydrolase; inositol transporter protein; integral membrane protein; Kds gene; KDH gene; KDH gene; KDH gene; KDH gene; ketone dehydrogenase (large subunit); ketone dehydrogenase (medium subunit); ketone dehydrogenase (small subunit); KtrA-like protein; LacI-family transcriptional regulator; LacI-family transcriptional regulatory protein; lipase (esterase); methyltetrahydrofolate cyclohydrolase; methyltetrahydrofolate dehydrogenase; MUR gene; MOA gene; MOA gene; MOA gene; MOA gene; MODA gene; MOD gene; MODC gene; MOA gene; molybdate-binding periplasmic protein modA; molybdenum transport ATPase modC; molybdenum transport transmembrane protein modB; molybdopterin cofactor synthesis protein moaA; molybdopterin cofactor synthesis protein moaB; molybdopterin cofactor synthesis protein moaC; molybdopterin synthase (large subunit moaE); multidrug resistance efflux protein; NAD-dependent aldehyde dehydrogenase; NADPH:quinone oxidoreductase; NDH gene; NDHM gene; NDHS gene; nicotine dehydrogenase (large subunit); nicotine dehydrogenase (medium subunit); nicotine dehydrogenase (small subunit); nitrile amino hydrolase; NrdH gene; OR0494 gene; ORF101; ORF103; ORF105; ORF106; ORF107; ORF110; ORF114; ORF116; ORF117; ORF119; ORF133; ORF137; ORF145; ORF146; ORF147; ORF153; ORF155; ORF16; ORF164; ORF165; ORF167; ORF168; ORF17; ORF170; ORF175; ORF176; ORF18; ORF180; ORF190; ORF192; ORF194; ORF196; ORF20; ORF201; ORF203; ORF204; ORF208; ORF21; ORF216; ORF218; ORF22; ORF221; ORF223; ORF224; ORF226; ORF231; ORF232; ORF236; ORF24; ORF241; ORF246; ORF249; ORF25; ORF26; ORF264; ORF27; ORF287; ORF29; ORF294; ORF30; ORF300; ORF302; ORF308; ORF309; ORF31; ORF310; ORF311; ORF314; ORF32; ORF323; ORF327; ORF33; ORF332; ORF339; ORF350; ORF357; ORF36; ORF367; ORF37; ORF38; ORF385; ORF388; ORF389; ORF395; ORF394; ORF402; ORF406; ORF407; ORF421; ORF424; ORF436; ORF458; ORF465; ORF470; ORF479; ORF48; ORF480; ORF481; ORF487; ORF497; ORF502; ORF550; ORF549; ORF583;

ORF594; ORF615; ORF64; ORF653; ORF666; ORF67; ORF68; ORF70; ORF72; ORF78; ORF82; ORF85; ORF87; ORF89; ORF890; ORF93; ORF96; ORF97; ORF99; oxidoreductase; PAR gene; permease; polyketide cyclase; PURU gene; RBA gene; recombinase; redoxin; ribose ABC transporter; RNA polymerase sigma factor; single-stranded DNA-binding protein; smf family protein; soj family protein; SSB gene; succinate-semialdehyde dehydrogenase; sucrose hydrolase; sugar ABC transporter; TBU gene; transcriptional regulator; transfer gene complex protein-like protein; transferase; transporter; transposase; WZT gene.

SOURCE
 ORGANISM
 Arthrobacter nicotinovorans
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Micrococciinae; Micrococcaceae; Arthrobacter.
 1 (bases 1 to 165137)
 Grether-Beck, S., Igloi, G.L., Pust, S., Schilz, E., Decker, K. and Brandsch, R.
 TITLE
 Structural analysis and molybdenum-dependent expression of the PA01-encoded nicotine dehydrogenase genes of Arthrobacter nicotinovorans
 MOL. Microbiol. 13 (5), 929-936 (1994)
 7815950
 95115562
 2 (bases 1 to 165137)
 Menendez, C., Igloi, G., Hemminger, H. and Brandsch, R.
 A PA01-encoded molybdopterin cofactor gene (moa) of Arthrobacter nicotinovorans: characterization and site-directed mutagenesis of the encoded protein
 Arch. Microbiol. 164 (2), 142-151 (1995)
 96172783
 8586735
 3 (bases 1 to 165137)
 Menendez, C., Igloi, G.L. and Brandsch, R.
 IS1473, a putative insertion sequence identified in the plasmid PA01 from Arthrobacter nicotinovorans: isolation, characterization, and distribution among Arthrobacter species
 Plasmid 37 (1), 35-41 (1997)
 97230479
 9073580
 4 (bases 1 to 165137)
 Menendez, C., Otto, A., Igloi, G., Nick, P., Brandsch, R., Schubach, B., Botcher, B. and Brandsch, R.
 Molybdate-uptake genes and molybdopterin-biosynthesis genes on a bacterial plasmid--characterization of MoaA as a filament-forming protein with adenosinetriphosphatase activity
 Eur. J. Biochem. 250 (2), 524-531 (1997)
 98088982
 9428706
 5 (bases 1 to 165137)
 Schenk, S., Hoelz, A., Krauss, B. and Decker, K.
 Gene structures and properties of enzymes of the plasmid-encoded nicotine catabolism of Arthrobacter nicotinovorans
 J. Mol. Biol. 284 (5), 1323-1339 (1998)
 99068870
 9878353
 6 (bases 1 to 165137)
 Baisch, D., Sandu, C., Brandsch, R. and Igloi, G.L.
 Gene cluster on PA01 of Arthrobacter nicotinovorans involved in degradation of the plant alkaloid nicotine: cloning, purification, and characterization of 2,6-dihydroxypyridine 3-hydroxylase
 J. Bacteriol. 183 (18), 5262-5267 (2001)
 21405725
 11514508
 7 (bases 1 to 165137)
 Igloi, G.L. and Brandsch, R.
 Sequence of the 165-kilobase Catalytic Plasmid PA01 from Arthrobacter nicotinovorans and identification of a PA01-dependent Nicotine Uptake System
 J. Bacteriol. 185 (6), 1976-1986 (2003)
 22505657
 12618462
 8 (bases 1 to 165137)
 Igloi, G.L. and Brandsch, R.
 REFERENCES
 AUTHORS

TITLE	Direct Submission
JOURNAL	Submitted (12-SEP-2002) Institute of Biochemistry and Molecular Biology, University of Freiburg, Hermann-Herder-Str. 7, Freiburg D-79104, GERMANY
FEATURES	Location/Oualifiers
source	1. .165137
CDS	/organism="Arthrobacter nicotianovorans" /mol_type="genomic DNA" /db_xref="taxon:29320" /plasmid="PAO1" complement(<1. .296) /note="ORF133"
CDS	/codon_start=1 /evidence=not_experimental /transl_table=11 /product="hypothetical protein" /protein_id="CAD47859.1" /db_xref="GI:25169023"
CDS	/translation="MIRTKATKKRRRROGAPGGTAEICDDGTGPNNRRKRGVDAPFGSPDEPPPTPKSTPGEDOCGRPHLPEPGLLVAAVVRRRROPAPPARRRTAGS" complement(344. .1138) /note="ORF264"
CDS	/codon_start=1 /evidence=not_experimental /transl_table=11 /product="putative hydrolase" /protein_id="CAD47860.1" /db_xref="GI:25169024"
gene	/translation="MPYTPSGIYWESGCGPYAAPAVILLIEGTGLIGMRDPCDLLAOGRLVLMDNRDGLSRREDGNIMADMADVIDYADADIDGKTITYGOSMGCLIAOHGALGPRMTWGLVFYTTPVLDIDPEGLITAEIPQHRDALASFEGDRATSPAMGAYEAMKRTELAGRMEFDAPDRSGLSRQRNAVALLDLPRPLTELTMVVALIHGRNDALIRARGSLRIAEVVPOELHLPGMGHEIAPALMDEFVAITRIAVC" complement(1140. .2093)
CDS	/gene="cSCSA" complement(1140. .2093) /gene="cSCSA"
CDS	/codon_start=1 /evidence=not_experimental /transl_table=11 /product="putative sucrose hydrolase" /protein_id="CAD47861.1" /db_xref="GI:25169025"
gene	/translation="MLAPEDRWIMDFEWVDHGIHLLYLLOAPKSLGDPELRNARNATTGHATSTLIDMTGHTGLRPQGGTVDAATWTGSVVRGDDLRMRYTGSAFLSPDSAGATERTAAVAVSSDLHHMAKDGFALADAWEKEGSDSSPEEAMMDPVYRREGLWMHLITARANTGEVMERGYYGHATSODLRMTARPRLTGEDGCAHLEIVQVSINQTPNIFLSHETVYAQAQRCDRAHNATGTMAAVTDRIELHANNTGPELYSGRVIAHNGLPVLLGLEFSGGGFTGGVSDPDVDELGLILKSGLSPORAG" complement(2124. .3092)
CDS	/gene="ESTra" complement(2124. .3092) /gene="ESTRa"
gene	/codon_start=1 /evidence=not_experimental /transl_table=11 /product="putative lipase (esterase)" /protein_id="CAD47862.1" /db_xref="GI:25169026"
CDS	/translation="MPYTRALQTLUDSLPATYGCIDLTDLPVAVROAGEANALAAVLAQEPLDVGSVHTAMPVNGEIKVKRIQQGTGPPRHIVFYHGGMAGTTEFFVDNVRCRRTALAIVTAVEERLARPERKFPILDSDSYALEMVEHADYDGDEENVITGGGSAGNSIAAGTAAIALAKARENENGHITFOYLEVPALDTLASPSVEYSGGEYPLPKNEIDLGVGYLRSPEDARNPYASPLLADDVSGLEPALIIISSEFPLDMDGSRXYARLDEVGVHAARLRGEGHHGSSQFKLLPEAAPAMDQYIDALKTHAADHRRVTP" complement(3148. .4482)
gene	/gene="OB0494" complement(3148. .4482) /gene="OB0494"
CDS	/codon_start=1 /evidence=not_experimental /transl_table=11 /product="putative inositol transport protein"

Db	34869	--CGCATGCGAGGCGATTAAGACCGCGCCACATGGTGCACACACTTGTACGCGAGATACCA	34926
Qy	294	ThrrgtgLeuAlaLeuThrrgllygTyrrgluLysIleAlaAlaIleAlaValArgsn	313
Db	34927	AAGGCTTTAACCC--ACGGGCTGCTGGAATCTCGGACCGAAGAACCCACGCTTACAA--	34980
Qy	314	LeuLeuAlaThrrArgsnProSerSerGlnLeuAspSerTyrIle-----	328
Db	34981	-----ATCACGCGTGGCCCTGACGGCGCGGCGAGGGGTGCTACACACACACCGTC	35034
Qy	329	-----MetAlaLysValTyrValLeuAlaSerGlyAlaIleGlysn	342
Db	35035	GGAACCTGCAGCGTCAACGGCGCTCCCTGTAGCTGTAGCAGGAAGCAATTCAGACCC	35094
Qy	343	ProGlnIleLeuTyrAsnSerGlyPheSerGlyLeuGlnValThrProArgsnAspSer	362
Db	35095	CTTACCGCTGCTCCCTGCTTGGCTACTCTGGGC-----CATACCAACGGC	35139
Qy	363	Leu-----	363
Db	35140	CTTGCCAACTCTCAGGCGAGTGGGGCGCACTATATGCGCACACACCGGTCTGTC	35199
Qy	364	-----IleProAsnLeuGlyAlaTyrIleThrGlnGlnProMetAlaPheGlysn	380
Db	35200	TACGCCAGATTCCCCCAACGAAATGCATATGTACCGCGCAACCAATGGACAGCATC--	35256
Qy	381	IleValLeuArgGlnIuPheValaAspSerValArgAspAspPro-----TyrGly	397
Db	35257	-----ATCACGCAAGAAATCGCGGCATGACCCATACAGGGGGCTTCGTT	35298
Qy	398	LeuProThrPrrPrrLysGluAlaValaIleGlnHisIleAlaLysAsnProThrAspAlaLeu	417
Db	35299	GCGGCTACATCATCGAGATGATCGCCAAAGGGCTCCCTCATTCAGTACC-----	35349
Qy	418	ProIleProPheArgAspProGluProGlnValThrThrProPheThrGlnIuHisPro	437
Db	35350	-----TTCAATGAGCCCG-----	35361
Qy	438	TrpHisThrGlnIleHisArgAspAlaPheSerTyrGlyAlaValGlyProGluValaAsp	457
Db	35362	-----GGGAGTGGGGCGCCACAAATTCACA	35385
Qy	458	SerArgValIleValAspLeuArg-----TrpPheGlyAlaThrAsp---Pro	472
Db	35386	GAAAGGTAAAGCTTACACCCCGCACAGCGGCACCTTGGATTTCGGGTGAGGACGTTCCG	35445
Qy	473	GluAlaAsnLeuLeuValaIlePheGlnAsnAspValaGlnAspGlyTyrSerMetProGln	492
Db	35446	CAGGCATCCAAAGAGAGACCTTAAACAACACGCTACGACGCGCTCGTCCGCCGA	35505
Qy	493	ProThrPheArgTyr--ArgProSer-----ThrAlaSerAsnValaArg	507
Db	35506	CCTGTGCTCCACTACGACGACATCCCAACAGACATCCGATCGGAGACCAACCGGTGAG	35565
Qy	507	IaArgLysMetAlaAspMetCysGluValaIaIaSerAsnLeuGlyGlyTyrLeuProT	527
Db	35566	CAGGCGCAGAAACTCTTCAAACTGTCCG-----	35594
Qy	527	hrSerProProGlnPheMetAspProGlyLeuAlaLeu---HisLeuAlaGlyThr	545
Db	35595	--AGCCATCCGCAACACAGAGCCCGGGGAGATGCCCTCGCGCCACAACCTGGGAACACCC	35652
Qy	546	ArgIleGlyPheAspLysAlaThrThrValaIaAspAsnSerLeuValIrrAspPhe	565
Db	35653	CGAATGACCAAGACCTGATGTAGAGCGGTCTCAACAGTTTGGGTCAAGCCACGACGTA	35712
Qy	566	AlaAsnLeuTyrValaIaGlyAsnGlyThrIleArgThrGlyPheGlyGluAsnProThr	585
Db	35713	CTTAACCTGTTGTCGACGCGGTCAACAATTACACACCGGGGCTGGCGGAAACCCACAC	35772
Qy	586	LeuThr 587	

DB	35773	CTGACA	35778
RESULT 14			
LOCUS	AE016780	305520 bp	DNA linear BCT 12-DEC-2002
DEFINITION	Pseudomonas putida KT2440 section 7 of 21 of the complete genome.		
ACCESSION	AE016780 AE015451		
VERSION	AE016780.1 GI:26557024		
KEYWORDS			
SOURCE			
ORGANISM	Pseudomonas putida KT2440		
REFERENCE	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas.		
AUTHORS	1 (bases 1 to 305520) Nelson,K., Paulsen,I., Weinel,C., Dodson,R., Hilbert,H., Fouts,D., Gill,S., Pop,M., Martins Dos Santos,V., Holmes,M., Brinkac,L., Beanan,M., DeBoy,R., Daugherty,S., Kolonay,J., Madupu,R., Nelson,M., White,O., Peterson,J., Khouri,H., Hance,I., Lee,P., Holtzapple,E., Scanlan,D., Tran,K., Moazzez,A., Ulteback,T., Rizzo,M., Lee,K., Kosack,D., Moestl,D., Wedler,H., Lauber,J., Hohelsel,J., Straetz,M., Helm,S., Kiewitz,C., Eisen,J., Timmis,K., Duesterhoft,A., Tummeler,B. and Fraser,C.		
TITLE	Complete genome sequence and comparative analysis of the metabolically versatile Pseudomonas putida KT2440		
JOURNAL	Environ. Microbiol. 4 (12), 799-808 (2002)		
REFERENCE	2 (bases 1 to 305520)		
AUTHORS	Nelson,K., Paulsen,I., Weinel,C., Dodson,R., Hilbert,H., Fouts,D., Gill,S., Pop,M., Martins Dos Santos,V., Holmes,M., Brinkac,L., Beanan,M., DeBoy,R., Daugherty,S., Kolonay,J., Madupu,R., Nelson,M., White,O., Peterson,J., Khouri,H., Hance,I., Lee,P., Holtzapple,E., Scanlan,D., Tran,K., Moazzez,A., Ulteback,T., Rizzo,M., Lee,K., Kosack,D., Moestl,D., Wedler,H., Lauber,J., Hohelsel,J., Straetz,M., Helm,S., Kiewitz,C., Eisen,J., Timmis,K., Duesterhoft,A., Tummeler,B. and Fraser,C.		
TITLE	Direct Submission		
JOURNAL	Submitted 05-NOV-2002 The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA		
FEATURES			
SOURCE	Location/Qualifiers 1..305520 /organism="Pseudomonas putida KT2440" /mol_type="genomic DNA" /strain="KT2440" /db_xref="taxon:160488" 108..1397 /gene="eno" /locus_tag="PP1612" 108..1397 /gene="eno" /locus_tag="PP1613" 1566..1847 /product="conserved hypothetical protein"		
gene			
CDS			
gene			
CDS			

gene
/protein_id="AA067234.1"
/db_xref="GI:24983095"
/translation="MRSPYMLFLVLLIGLIGLOXRIMWNGSLAQVTELEKQOIQDQHA
ENERLERNRYLDAEVLLEKKMETVEERANHELGWKEGTELLQLRQK"
1844..2551
/gene="ispd"
/locus_tag="pp1614"
1844..2551
/gene="ispd"
/locus_tag="pp1614"
/note="Identified by match to TIGR protein family HMM
TIGR00453"
/codon_start=1
/transl_table=1
/product="4-diphosphocytidylyl-2C-methyl-D-erythritol
synthase"
/protein_id="AA067235.1"
/db_xref="GI:24983096"
/translation="MIDILPFAWAVIPAPAGVARMADRPKOYLELAGOULLEHSLDC
FLGHPALKGVVSTAEDDPYWPLRGASDPRIHCAAGGERADSVLNAALVLAHQGA
DSQWVLYHDAAPMLARSDDLKILSELADDPVGLLAVPADTLKRRADSHGVSATVD
RSTIQAATYTPQMFRLGALHRALECLVSDVYVDEASAIEMSGQAPRLVEGRSDNIKY
TRPDLEMLRQRMAGKR"
complement(2679..3575)
/locus_tag="pp1615"
complement(2679..3575)
/locus_tag="pp1615"
/note="Identified by match to PFM protein family HMM
PF03466"
/codon_start=1
/transl_table=1
/product="transcriptional regulator, lysr family"
/protein_id="AA067236.1"
/db_xref="GI:24983097"
/translation="MSSRWEGIDEFVAVESGQPTAAERLGVSSSHISQIARLEDR
LQTRLYRSTRRYTLSEAGQFLQHCRLQDREARLAMDGLASEPGLRLMTCAVA
YGERFIPLVTRFMALYPOLRVDELNHTLDLHEGMLAIRGLRSLDVLATRIA
PRMYLCAVPAYLERYGRPHSLSELARNHNCVSSDMLAQDQREISORQGWRCN
SGQAVLAAALOGIGLCOLPDYVLEHNSGALVLELHNPNTAAVVALTYQQRHLSF
KVRRLVYLKEGLAGLPETRRG"
3680..4795
/locus_tag="pp1616"
3680..4795
/locus_tag="pp1616"
/note="Identified by match to TIGR protein family HMM
TIGR01751"
/codon_start=1
/transl_table=1
/product="D-isomer specific 2-hydroxyacid dehydrogenase
family protein"
/protein_id="AA067237.1"
/db_xref="GI:24983098"
/translation="MITSRAAFAFEAKKPLEIVEVDVAMRKAGEVLLRVASGCVHT
DAYTLGSADEPGIIPSLIHGGAIVAVGVSILAVGDHVIPLYPECGCKFCIS
GKTMICQAIRATQGLMDGTFRYSKGOOLFMYMTSTFSEYTLPEISVAKIOKE
APLEKVCILGCGVTTIGAVLNTAKVKGPDVALFEGGIGLSAIVGAVKAKRIIA
IDINPAKFETAIROLGATDCAVMPKXDRIQDVYDLDDGVDSPFECTGNQIIMBAL
ECCCHKNGESIITIGVAGAGEISTRPOLYTGKRWGRSARFGVGRSELPSYVMSER
GEPLDTFITHMGLDINKAFDLMEGKSTRSVTH"
4804..5658
/locus_tag="pp1617"
4804..5658
/locus_tag="pp1617"
/note="Identified by match to PFM protein family HMM
PF00756"
/codon_start=1
/transl_table=1
/product="esterase, putative"
/protein_id="AA067238.1"
/db_xref="GI:24983099"
/translation="MSLDNISQKSEFGQWKKRYRHHSKVLGCDVAVPVLPEPQAGEE
KLPLVWLSGITCTDENFMQAGQALAAELGLITVADPSPREQVPGDPDGMDFG
LGAGFYLNATQOPWAQHYRMHDVVEELPALIEAHFPASGRSISGSHSGHGLAYCA

gene
/protein_id="AA067239.1"
/db_xref="GI:24983100"
/translation="MRIGHGQVHRRPDDGPTITGVRIPKHYGLLAHSDQVLLHAL
SDALLGALADSDIGKHPDDDPQKGRDRLNLHVYGYKAKKWKGNVDTATVAAQ
PKMAPIHETMRQLAIEDLQVELDQVNVKATTEKLGFTGREGIIVASHVALLPA"
6241..7299
/locus_tag="pp1619"
6241..7299
/locus_tag="pp1619"
/note="Identified by match to TIGR protein family HMM
TIGR00094"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AA067240.1"
/db_xref="GI:24983102"
/translation="MTELELLGPRASEPPLGTAIILKAVAEDQVDEVILPISGQGEH
LMLVERDLNTEEARLARAAGVPAVSISYAGIKROALTQWFSIHLPGKADPL
SRADASLRVLYKQVHORLQORGAHSAANGFTLRLLTALAADHQAVDARLEOLRQGVN
YFGTORHGGGVNDALDMARKALPEORVRVRSILSAGRSYVFNOLLAHVADGSW
ARAOVDILATDTSRSPRACACADPRILIIDLHPGPMWGCASAPAGAPOLEN
AIGARQPLCQWLAQAGMDHERILRLPTGGLTWHTPEPDLQLEFVLPACCFATVY
REVYDLVSAQGTDSPCVF"
7287..8036
/gene="sure"
/locus_tag="pp1620"
7287..8036
/locus_tag="pp1620"
/note="similar to GP:8141682; identified by sequence
similarity: putative"
/codon_start=1
/transl_table=1
/product="stationary-phase survival protein Sure"
/protein_id="AA067241.1"
/db_xref="GI:24983103"
/translation="MRLISNDGCVTAPGIALHAALADYAECAVIAAPDDKSGASS
LTIDRPLHPOTLANGFISINCTPTDCVHLGLNGLLPQMPDQVVGINLGAHGLDVLV
SGTYAALLEGFLGCTSLARSLRSROPNLPTAATYARLVEAOSRLVLPRTYLVNIN
IPNLPLEHIRIGIQLTRUGHRARAAPTKVAVPVRKKEGVIWAVAGDAEDGGGTDFHNV
MGGVSTPPLDRLTFDAFQDGLMGLGL"
8015..8710
/gene="pcm"
/locus_tag="pp1621"
8015..8710
/gene="pcm"

Alignment Scores:
Pred. No.: 3,41e-05 Length: 305520
Score: 227.00 Matches: 133
Percent Similarity: 32.39% Conservative: 95
Best Local Similarity: 18.89% Mismatches: 264
Query Match: 6.91% Indels: 212
DB: 1 Gaps: 27

US-09-856-327-2 (1-618) x AE016780 (1-305520)

OY 29 GIUthrTyrgIyaAnaspGlyValAspValPheIlealagIysergIyProIleGlyAla 48
DB 47992 CAAACCGAGAACATCGAAACCGCGATGTGATCGTGTGGCGCGCGATGCGCGAGT 48051

QY	49	ThrTyrAlaLysLeuCyValGlnAlaGlyLeuAlaGlyValAlaMetValGlnIleGlyAla	68
Db	48052	GTCAATGCTTACCAAGCTGGTGGGCTGGCCGGGTGAAAGTGTGGTCTCAATCGGCG	48108
QY	69	AlaAspSerPheTyrAlaValAlaAsnAlaGlnGluGlyThrAlaValProGlyValProGly	88
Db	48109	-----	48120
QY	89	TyrHisLysLysAsnGluIleGluPheGlnLysAspIleAspArgPheValAsnValIle	108
Db	48121	-----	48162
QY	109	LysGlyAlaLeuGlnGlnValSerValProValArgAsnGlnAsnValProThr	126
Db	48163	CTCAAGCGCGCCGAGTCCGCGGTACCCACCTGTAGCAGACGACGAGACCCGGCCAG	48222
QY	127	-----	48282
Db	48223	TTCCGGCCACGGCGGCACCATCCAGACCTCATCTCACTCAAGTACCCAAAGCCCAACAG	48282
QY	137	GlySerSerAlaIleSerAsnGlyLysAsnProHisGlnArgGluPheGlnLysLeuSer	156
Db	48283	CGCCAGAAAGCTCACTCGTGCAGAAAGGCGCCAC	48324
QY	157	AlaGlnAlaValThrArgGlyValGlyGlyMetSerThrHisThrPheCysSerThrPro	176
Db	48325	GCCAGCACTTACGACGGGTAGGGGGGCGCCACCATGTGCAGTGCAGTGGGCGCCGCTG	48384
QY	177	ArgIleHisProPheMetGluSerLeuProGlyIleGlyArgProLysLeuSerAsnAsp	196
Db	48385	CGCGTG-----	48426
QY	197	ProAlaGluAspAspLysGluThr-----	209
Db	48427	GTCCGGGTGCAC-----	48477
QY	210	GluLysGluArgLeuIleGlyLysSerThrLys-----	220
Db	48478	CGCCCGAAGAGAGATGGGTGTCCGCCGATGTGCCTCGACACTTACCTGGGCGTG	48537
QY	221	GluPheAspGluSerIleArgHisThrLeu-----	237
Db	48538	ACGTTCCCGCAGGGCTCAATACAGATCATGGGATCCCGCGCTGGTGTGCACAG	48597
QY	238	-----	246
Db	48598	GGCCTGGGCAAGCACTCACGGGACCGGTATACCAAGCCCGGAGAGGGCGCTGACGCG	48657
QY	247	ArgProLeuProLeuAlaCysHisArgLeuLysAlaValProGluTyrValGluThrPhe	266
Db	48658	AAGCGGTACCCAGGTGTCACTCAATGTCCGGCAAAACCCCG	48699
QY	267	SerIaGluAsnLeuPheHisSerIleTyrAsnAspAspLysGlnLysLysLeuPheThr	286
Db	48700	-----	48738
QY	287	LeuLeuThrAsnHisArgCysThrArgLeuAla-----	303
Db	48739	TGCCTGGCAACACCACTGCACGGCGGATGTGCCGATCGAAGGCCAAGTACGACGGCAC	48798
QY	304	-----	317
Db	48799	GTGACCATGGCCAAAGGCACTGGATTACCGGCAAGGTCCGATCTTACACAGACCGTTGCA	48858
QY	318	-----	319
Db	48859	AGCCAACTGCAGTGCATGCCAAATGGGGTACGCGGCATTCGATTCATGACCAAGAC	48918
QY	320	ProSerSerGlnLeuAspSerTyrIleMetAlaLysValTyrValLeuAlaSerGlyAla	339
Db	48919	AATGTATCGGCGCAAGACGGCAGCGCCCAAGGCAACGCTACTGTATTCGCCGCCACGCG	48978
QY	340	IleGlyAsnProGlnIleLeuTyrAsnSerGlyPheSerCylLeuGlnValThrProArg	359

Db	48979	ATCGAATCACCAGAGCTGCTGCTCACTCC-----	ATCGGCCCGGAAC	49020
QY	360	AsnaSperLeuIIleProAsnLeuGlyArgTyrIleThrfGluInProMetaIaphcys	379	
Db	49021	AGCCCTAAGGCGTGGCCAACACGACGAGGCGCAGGTGCGCCAG-----	49065	
QY	380	GlnIleValLeuArgGlnGluProHeValAspSerValArgAspAspProTyrGlyLeuPro	399	
Db	49063	-----GCATTGCTGACCCAGCGGCTGTACTCGGCC	49092	
QY	400	TrpTrpIysGluAlaValAlaGlnHisIleAlaLysAspProThrAspIaleuProIle	419	
Db	49093	TGG-----GCCCTGATGCGCCAGAGCAAGCCGCTGTC	49125	
QY	420	ProPheArgAspPro-----GluProGlnValThrTrpPhe---	432	
Db	49126	CCGTTCCGGGGCGCGTGTCTACTCGGGGATGAGACATGGCGCATGGCCGTTCCGC	49185	
QY	433	ThrGlnGluHisProThrPheIsthrGlnIleHisArgAspIaphSerTyrGlyAlaVal	452	
Db	49186	AGCCAGCGCTGCGCCGCTGCGATGCAATCGGCACACGAGGCGTGAACCTGGCTGAAC	49245	
QY	453	GlyProGluValAspSerArgValIleValAspLeuArgTrpPheGlyAlaThrAspPro	472	
Db	49246	GACCCCTACACCTTCGCTGTGATTTCATGACGGCACCAACACGAGCGGCGCTAACCCC	49305	
QY	473	-----GluAlaAsnAsnLeuValPhe	480	
Db	49306	AGCAATCTGGCGGTGTCCGGCACAGGCGCTGTGCACAGCTCAACAGCTGGTGCACCGC	49365	
QY	481	Gln-----	481	
Db	49366	CAGTTCCGCTGGCTTCTTGTCGACAGGTGAGAAAGCCCGATATGCGCTGTGC	49425	
QY	482	-----AsnaSperValGlnAspGlyTyrSerMetProGlnProThrPheArg	496	
Db	49426	CGGTGAGAGAGATGACGAGATTCTCCGACGGATTGGCGCTGCTCGGCCGACGATCAC	49485	
QY	497	TyrArgProSer-----ThrAlaSerAsnValArgAlaArgLysMetMetaIaaspMet	514	
Db	49486	TACGAACTGTCCGACTACACCAAGAGAGGTTCAGACAGGCGGCGCTGACTATGAC	49545	
QY	515	CysGluValAlaSerAsnLeuGlyGlyTyrLeuProThrSerProGlnPheMetAsp	534	
Db	49546	ATC---ATCACCGAGCTGATGGGCGCTACCGAGCTGACCGACTGAACCTTAAGCTCGAG	49602	
QY	535	ProGlyLeu-----AlaLeu	539	
Db	49603	CCCCGCTCGCAACCGGTGTTGAGTACGACGCGCCAGACTACCGCTTACGGCGCTGCG	49662	
QY	540	HisLeuAlaGlyThrThrArgIleGlyPheAspLysAlaThrThrValAlaAspAsnAsn	559	
Db	49663	CACCTGAATGGCGACCTTACCGCATGGCGTCCGACCGCACCAAGTCGGTGTGGACAAAGAC	49722	
QY	560	SerLeuValTrpAspPheAlaAsnLeuTyrValAlaGlyAsnGlyThrIleArgTyrGly	579	
Db	49723	CAGGCGAGCTGACACCCCAACCTGTCTCTGTGGCGACGCGCTGTCCCACTTACC	49782	
QY	580	PheGlyGluAsnProThrIleuThrIleuThrSerMetCysHisAlaIleLysSerAlaArgSerIle	599	
Db	49783	GGCACCGCGCAACCGGACCTTGACATACACCGCGTGAAGTTCACAGCGCCGACAGCTGTG	49842	
QY	600	IleAsnThrLeu	603	
Db	49843	GCCAGCGACCTTG	49854	
RESULT 15				
LOCUS	AF430844	2465 bp	DNA	linear
DEFINITION	Burkholderia cepacia glucose dehydrogenase (gdhAlpna) gene,			
ACCESSION	AF430844			

VERSION AF430844.1 GI:23820937
KEYWORDS Burkholderia cepacia
SOURCE Burkholderia cepacia
ORGANISM Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales; Burkholderiaceae; Burkholderia; Burkholderia cepacia complex.
REFERENCE 1 (bases 1 to 2465)
AUTHORS Inose, K., Fujikawa, M., Yamazaki, T., Kojima, K. and Sode, K.
TITLE Cloning and expression of the gene encoding catalytic subunit of thermostable glucose dehydrogenase from Burkholderia cepacia in Escherichia coli
JOURNAL Biochem. Acta 1645 (2), 133-138 (2003)
MEDLINE 22461202
PUBMED 12573242
REFERENCE 2 (bases 1 to 2465)
AUTHORS Sode, K., Inose, K., Fujikawa, M. and Yamazaki, T.
TITLE Direct Submission
JOURNAL Submitted (09-OCT-2001) Biotechnology, Tokyo University of Agriculture and Technology, 2-24-16 Naka-machi, Koganei, Tokyo 184-8588, Japan
FEATURES
source Location/Qualifiers
1..2465
/organism="Burkholderia cepacia"
/mol_type="genomic DNA"
/db_xref="taxon:292"
762..2381
/gene="gdhAlpha"
762..2381
/gene="gdhAlpha"
/codon_start=1
/transl_table=1
/product="glucose dehydrogenase"
/protein_id="FAN39686.1"
/db_xref="GI:23820938"
/translation="MADTDITOKADVIVVGVSGVAGIAYAHOLAMAGKAVILLIENRPM
RWELVERNQPKMDPMAPYSSPMAPHEGYDPNDYLILKGHKNSOYIRAVGT
TWHMAASMRPIPNDFEMKSYGVGRMPLOYDLEPYORAEELGVCGPREEDY
SPKOPYMPPLPLSFNEOTIKTALNNDKPHVTEPVARSNPYGRPCGNNC
MPLCPICANGIIVHEKAEKAGAKLLENVYVLEKLEPRKRIYALYKKTGAENR
EGRYEVLANGIETPKLILASNDPENGANSNDMGRLMDHPRGVSFYAEKIM
PGKRGOMTLIGFRDGFPRATEAKKIHLSNRSRIDQETOKIFKAKLMPDELDO
IKRDSARYVDFCHETLPPENRIVSKTATDAGIIPREITYAIDDYVKGAAHR
EYATATAKVLGGDNDVNFDEAFPNNHITGSTIMGADARDSVDKCRTPFHPNLFISS
SATMPYGVGNVTLTIALALRMSDTLKREY"
BASE COUNT 497 a 792 c 769 g 407 t
ORIGIN
Alignment Scores:
Pred. No.: 1..21e-06 Length: 2465
Score: 208.50 Matches: 131
Percent Similarity: 34.05% Conservative: 91
Best Local Similarity: 20.09% Mismatches: 221
Query Match: 6.35% Gaps: 209
DB: 1 Indels: 29
US-09-856-327-2 (1-618) x AF430844 (1-2465)
QY 37 AspvAlpheIleAlaGlySerGlyProIleGlyAlaThrTyraAlaLysLeuGlyValGlu 56
DB 789 GACGTGCTGCTGCTGGATCGGTCGCGGGCCGATCGTCGGCATCAGCTCGCGATG 848
QY 57 AlAGlyLeuArGValValMeValGluIleGlyAlaAlaAspSerPheTyraAlaValAsn 76
DB 849 GCGGGCAAGCGCGATCTGCTCGAAGCGGC----- 881
QY 77 AlAGluGluGlyThrAlaValProGlyValAlaProGlyTyraHisLysLysAsnGluIleGlu 96
DB 882 -----CCGCGCATCGCGCGCTG-----GAATC--- 905
QY 97 PheGlnLysAspIleAspArgPheValAsnValIleLysGlyAlaLeuGlnGlnValSer 116
DB 906 -----GTGAGCGCTTCGCAAT----- 923

QY 117 ValProValArgAsnGlnAsnValProThrLeuAspProGlyAlaTrpSer----- 133
DB 924 CAGCCCGCAGAGGAGTTCATGCGCGCCGATACCGCGTACCCCGTGGCGCCGATGCC 983
QY 134 -----AlaProGlySerSerAlaIleSerAsnGlyLysAsnProHisGlnArgGlu 151
DB 984 GAGTACGCGCCGCGCAAGCAGATACCTGATCTCGAAGGCGGACAGCAAGTTCACGCGAG 1043
QY 152 PheGlnLysLeuSerAlaGluAlaValThrArgGlyValGlyGlyMetSerThrHisTrp 171
DB 1044 TAC-----ATCGCGCGGCGGCGGCGACGACGATGCGACATGG 1079
QY 172 ThrCysSerThrProArgIleHisPro-----PrometGluSerLeuProGlyIle 188
DB 1080 GCGCGTGGCGGTGGCGCTTCATTCGACAGCATCTCAAGATGAAAGACGCTAGCGGCTC 1139
QY 189 GlyArgProLysLeuSerAsnAspProAlaGlu---AspAspLysGluTrpAsnGluLeu 207
DB 1140 GCGCGC-----GACTGCGCATCCAGTACGACGATCTCGAG-----CCGTAC 1181
QY 208 TyrSerGluAlaGluArgLeuIleGlyThrSerThrLysGluPheAspGluSerIleArg 227
DB 1182 TATCAGCGCGCGGAGAGAGACCTGCGGTGGCGCCGCGCGCGAGAA----- 1232
QY 228 HisThrLeuValLeuArgSerLeuGlnAspAlaTrpLysAspArgGlnArgIlePhe--- 246
DB 1233 -----GATCTGATCTCGCGCGCGCAAGCAGCCGATGCCG 1265
QY 247 ---ArgProLeuProLeuAlaCysHisArg-----LeuLysAsnAlaProGluTrpVal 263
DB 1266 ATCGCGCGCGTGGCGGTGTGCTGTCGACAGCAGACATCAAGACGCG----- 1313
QY 264 GluTrpHisSerAlaGluAsnLeuPheHisSerIle-----TyrAsnAsp 278
DB 1314 ---CTGACCAACTAGACATCCGAGATCCATCTCTGTGACCGAGCCGCGTGGCGCGCAAGC 1370
QY 279 AspLysGlnLysLysLeuPheThrLeuLeuThrAsnHisArgCysThrArgLeuAlaLeu 298
DB 1371 CGCCCGTACAGCGCGCGCCCGACACTTGTGGCGCAACAACAAGTCAATCGCATGTGCCG 1430
QY 299 ThrGlyGlyTyr-----GluLysLysIleGlyAlaAla 309
DB 1431 ATCGCGCGATGTACACAGCGCATGTCACGTCGAGAGAGCGGACGCGCGCGCGAAG 1490
QY 310 GluValArgAsnLeuLeuAlaIleThrArg-----AsnProSerSerGlnLeu----- 324
DB 1491 CTGATCGAGACGCGGTGCTGTACAGCTGAGACGCGGCCGCAACAGCGATGTCGCG 1550
QY 325 -----AspSerTyrIleMetAlaLysValTyrValLeu 335
DB 1551 GCGCTTCACAAAGACAGACAGCGCGCGGCGGATCGCTCGAAGCAAGTATTCGTGCTC 1610
QY 336 AlaSerGlyAlaIleGlyAsnProGlnIleLeuTyrAsnSerGlyPheSerGlyLeuGln 355
DB 1611 GCCCGCAAGCGATCGACGCGCGCAAGATCTCTG----- 1643
QY 356 ValThrProArgAsnAspSerLeuIleProAsnLeuGlyArgTyrIleThrGluGlnPro 375
DB 1643 ----- 1643
QY 376 MetAlaPheCysGlnIleValLeuArgGlnGluPheValAspSerValArgAspAspPro 395
DB 1644 -----CTGATGTCGCGAAGCCGCGATTTCCG 1670
QY 396 TyrGlyLeuProTrpTrpLysGluAlaValAlaGlnHisIleAlaLysAsnProThrAsp 415
DB 1671 AACGCTGTGCGAAGACAGCTGACATGTGCGCGCAACCTGATGAGACATCCCGGCAACC 1730
QY 416 AlaLeuProIle-----ProPheArgAspProGluProGlnVal 428
DB 1731 GCGCTGTCGTTTATGCGAGCGAGACCTGTGGCGGGCGGCGCGCGAGACATGACG 1790
QY 429 ThrThrProPheThrGlnGluHisProTrpHisThr-----GlnIleHis 443

```
Db      1791 TCCTGATCGTTTCGGACGCTCCGTCGCCGCCGACCGAAGCGCGAAGAGATCCAC 1850
QY      444 -----ArgAspAlaPheSerTyrGlyAlaVal 452
Db      1851 CTGTGCAACCTGTCCGCGATCGACCGAGACGACGACAGATCTTCAAGGCCGCAAGCTG 1910
QY      453 -----GlyProGluValAspSerArgVal-----IleValAspLeu 464
Db      1911 ATGAAGCCCGACGAGCTCGACCGAGATCCGACCGCTCCGACGCTACGTCGAGTTC 1970
QY      465 ArgTyrPheGlyAlaThrAspProGluAlaAsnAsnLeuValPheGlnAsnAspVal 484
Db      1971 GACTGCTTCACGAAATCTCGCCGCAACCGCAACCCGATCGTCCCGAGCAAGACGGCG 2030
QY      485 GlnAspGlyTyrSerMetProGlnProThrPheArgTyrArgProSerThrAlaSerAsn 504
Db      2031 ACCGATGCGATCGGCATTCGCGCGCCCGCGATCAGCTAT-----GCGATCGAC 2078
QY      505 ValArgAlaArgLysMetMetAlaAspMetCysGluVal-----AlaSerAsn 520
Db      2079 GACTACGTGAAGCGCGCGCGCGCATACGCGAGGTCTACGCGACCGCGCGAAGCTG 2138
QY      521 LeuGlyGly-----TyrLeuProThrSerProProGlnPhe 532
Db      2139 CTCGGCGGCGACGACGCTCGTTCAACGACGAATTGCGCGCGACACAT----- 2186
QY      533 MetAspProGlyLeuAlaLeuHisLeuAlaGlyThrThrArgIleGlyPheAspLysAla 552
Db      2187 -----CACATCACGGGCTCGACGATCATGGCGCGCATGCGCGC 2225
QY      553 ThrThrValAlaAspAsnAsnSerLeuValTyrAspPheAlaAsnLeuTyrValAlaGly 572
Db      2226 GACTCGTCGTGACACAGGACTGCCGACGCTTCGACCATCGAACCCTGTCATTTCGAGC 2285
QY      573 AsnGlyThrIleArgThrGlyPheGlyGluAsnProThrLeuThrSerMetCysHisAla 592
Db      2286 ACCGGCAGCATGCCGACGCTCGTACCGTAACGTAACGTGACGCTGACGATCGCGCGCTCGCG 2345
QY      593 IleLysSerAlaArgSerIleIleAsnThrLeuLys 604
Db      2346 CTGCGGATGTCG-----GACACGCTGAAG 2369
```

Search completed: September 24, 2003, 07:57:18
Job time : 6746 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 24, 2003, 01:45:38 ; Search time 7651 Seconds
(without alignments)
11260.708 Million cell updates/sec

Title: US-09-856-327-1
Perfect score: 2106
Sequence: 1 atcagcgcattctctctca.....aaaaaaaaaaaaaaaaaa 2106

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

GenBankl: *
1: gb_ba: *
2: gb_htg: *
3: gb_in: *
4: gb_cm: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sts: *
12: gb_sy: *
13: gb_un: *
14: gb_vl: *
15: em_ba: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_mu: *
20: em_or: *
21: em_ov: *
22: em_ov: *
23: em_pat: *
24: em_ph: *
25: em_pl: *
26: em_ro: *
27: em_sts: *
28: em_un: *
29: em_vl: *
30: em_htg_hum: *
31: em_htg_inv: *
32: em_htg_other: *
33: em_htg_mus: *
34: em_htg_pln: *
35: em_htg_rod: *
36: em_htg_mam: *
37: em_htg_vrt: *
38: em_sy: *
39: em_hgt_hum: *
40: em_hgt_mus: *
41: em_hgtg_other: *

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2106	100.0	2106	6 BD013176	BD013176 A novel p
2	225.8	10.7	2069	8 AY291124	AY291124 Trimetes
3	221	10.5	2064	8 AF535193	AF535193 Peniphor
4	212.4	10.1	1869	6 AR141573	AR141573 Sequence
5	210.2	10.0	1869	6 E11766	E11766 cDNA of pyr
6	210.2	10.0	1869	6 I82410	I82410 Sequence 1
7	210.2	10.0	2046	8 D73369	D73369 Coriolus ve
8	175.6	8.3	1902	6 E32737	E32737 Panus cornu
9	152.6	7.2	1701	6 AR170485	AR170485 Sequence
10	129	6.1	1946	6 E33665	E33665 Novel prote
11	129	6.1	1946	6 E37379	E37379 Reagent for
12	129	6.1	1946	8 AB043883	AB043883 Tricholom
13	94.6	4.5	2013	9 AB070201	AB070201 Macaca fa
14	94.4	4.5	589	10 BC049706	BC049706 Mus muscu
15	94.4	4.5	1901	5 BC048876	BC048876 Danilo rer
16	94.4	4.5	4640	6 AX376008	AX376008 Sequence
17	93.8	4.4	3005	10 BC032271	BC032271 Mus muscu
18	92	4.4	2360	10 BC008547	BC008547 Mus muscu
19	92	4.4	2578	9 AB049758	AB049758 Homo sapi
20	92	4.4	3660	10 BC051069	BC051069 Mus muscu
21	91.6	4.3	1618	3 AY069072	AY069072 Drosophil
22	91.6	4.3	1624	3 AK114250	AK114250 Clona int
23	91.4	4.3	1911	10 BC048954	BC048954 Mus muscu
24	91.2	4.3	2091	6 BD176856	BD176856 A method
25	91.2	4.3	2091	17 AF111849	AF111849 Homo sapi
26	91.2	4.3	2964	6 AX574582	AX574582 Sequence
27	91	4.3	1282	9 BC017745	BC017745 Homo sapi
28	90.8	4.3	845	9 BC044653	BC044653 Homo sapi
29	90.8	4.3	1327	6 AX306545	AX306545 Sequence
30	90.8	4.3	1817	10 BC051174	BC051174 Mus muscu
31	90.6	4.3	1567	9 AB047887	AB047887 Macaca fa
32	90.6	4.3	2537	9 BC003684	BC003684 Homo sapi
33	90.4	4.3	2335	9 AB055368	AB055368 Macaca fa
34	90.4	4.3	3983	9 HSM803572	HSM803572 Homo sapi
35	90	4.3	1289	9 BC032350	BC032350 Homo sapi
36	90	4.3	2384	3 AY069709	AY069709 Drosophil
37	90	4.3	2566	17 AF130082	AF130082 Homo sapi
38	90	4.3	3809	9 HSM804486	HSM804486 Homo sapi
39	89.8	4.3	470	3 AF177908	AF177908 Hydra vul
40	89.8	4.3	1864	9 AB056421	AB056421 Macaca fa
41	89.8	4.3	2180	8 AB028665	AB028665 Gentiana
42	89.8	4.3	4861	9 BC040342	BC040342 Homo sapi
43	89.6	4.3	1286	9 BC044234	BC044234 Homo sapi
44	89.6	4.3	2032	9 AK027204	AK027204 Homo sapi
45	89.6	4.3	2260	6 BD021751	BD021751 Novel gen

ALIGNMENTS

RESULT 1
BD013176
LOCUS
DEFINITION A novel protein, a gene coding therefor and a method of using the same.
ACCESSION BD013176
VERSION BD013176.1 GI:22093365
KEYWORDS
SOURCE WO 0121657-A/1.
ORGANISM Lyophyllum shimeji
Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
Agaricales; Tricholomataceae; Lyophyllum.
REFERENCE 1 (bases 1 to 2106)
AUTHORS Takakura,Y., Kuwata,S. and Inoue,Y.
TITLE A novel protein, a gene coding therefor and a method of using the

Pred. No. is the number of results predicted by chance to have a

JOURNAL Patent: WO 0121657-A 1 29-MAR-2001;
JAPAN TOBACCO INC., CORPORATE JURIDICAL PERSON SOCIETY FOR TECHN
INNOVATION OF AGRICULTURE YUKIO SHIMOMURA, SHIRO TAKEKAWA, NOBIO CHO
FORESTRY AND FISHERIES, YOSHIMITSU TAKAKURA, SHIGERU KUMATA, YASUHIRO
INOUE
OS Lyophyllum shimeji
PN WO 0121657-A/1
PD 29-MAR-2001
PF 20-SEP-2000 WO 2000JP006404
PR 21-SEP-1999 JP 99P 267238
PI YOSHIMITSU TAKAKURA, SHIGERU KUMATA, YASUHIRO INOUE PC
C07K14/375, C12N15/31, C12N15/63, C12N1/21, C12Q1/68, C12P21/02, PC
A01M65/00
CC
FH Key Location/Qualifiers
FT CDS (8)..(1861).
FEATURES
source 1..2106 location/Qualifiers
BASE COUNT 593 a 540 c 527 g 446 t
ORIGIN
Query Match 100.0%; Score 2106; DB 6; Length 2106;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATCAGCATGCTCTCTCAACGACGAGATGCTAGCGGCTATCCAGCGTCTATGCAAT 60
DB 1 ATCAGCATGCTCTCTCTCAACGACGAGATGCTAGCGGCTATCCAGCGTCTATGCAAT 60
QY 61 CAACGACAGATTCTTAAGAACGCAATTCAGAAACATACGGAACGAGAGATTGAT 120
DB 61 CAACGACAGATTCTTAAGAACGCAATTCAGAAACATACGGAACGAGAGATTGAT 120
QY 121 ATTCAATTGACAGATCTGGACCCATTGGAGCGACGATATGCAAAAGCTCTGTGTTGAAGCTGG 180
DB 121 ATTCAATTGACAGATCTGGACCCATTGGAGCGACGATATGCAAAAGCTCTGTGTTGAAGCTGG 180
QY 181 TCTAGTGTGTGATGATGATGAGATGAGACCTGCTGATGCTCTAGCGCTTAAATGCGGA 240
DB 181 TCTAGTGTGTGATGATGATGAGATGAGACCTGCTGATGCTCTAGCGCTTAAATGCGGA 240
QY 181 TCTAGTGTGTGATGATGATGAGATGAGACCTGCTGATGCTCTAGCGCTTAAATGCGGA 240
DB 181 TCTAGTGTGTGATGATGATGAGATGAGACCTGCTGATGCTCTAGCGCTTAAATGCGGA 240
QY 241 AGAAGAACTGCAAGTTCCCTACGTTCTGCGTACCAACAAGAAATGAATGAGATTCCA 300
DB 241 AGAAGAACTGCAAGTTCCCTACGTTCTGCGTACCAACAAGAAATGAATGAGATTCCA 300
QY 301 GAAAGATATATGACCGCTTCTGCAATGTATCAAGGAGAGCCTTACAAAGATCTCTGTTCC 360
DB 301 GAAAGATATATGACCGCTTCTGCAATGTATCAAGGAGAGCCTTACAAAGATCTCTGTTCC 360
QY 361 TGTCAAGAAACGAAACGATGCTACACTTGATCCCGAGGCTTGAGCGGCCCTTGGAAG 420
DB 361 TGTCAAGAAACGAAACGATGCTACACTTGATCCCGAGGCTTGAGCGGCCCTTGGAAG 420
QY 421 TTCAGCCATATCGAAACGCTTAAATCTCCACAGCGGGAATTCGAAACTTGTCTGCGGA 480
DB 421 TTCAGCCATATCGAAACGCTTAAATCTCCACAGCGGGAATTCGAAACTTGTCTGCGGA 480
QY 481 GGGCCCTAACGCGTGGAGTCGGCGCATGAGTACCCACTGTGAGCTCTCCACCCAGGAT 540
DB 481 GGGCCCTAACGCGTGGAGTCGGCGCATGAGTACCCACTGTGAGCTCTCCACCCAGGAT 540
QY 541 TCATCCAGCCATGAAAGTCTCCGGGATCGCGGCTCGGAAGCTCAGTAACGACCCGGC 600
DB 541 TCATCCAGCCATGAAAGTCTCCGGGATCGCGGCTCGGAAGCTCAGTAACGACCCGGC 600
QY 601 AGAGAGACAAAGAGTGAAGAGAGCTTTATTCGAGGCGGAGCGCTCATCGGAGACTTC 660
DB 601 AGAGAGACAAAGAGTGAAGAGAGCTTTATTCGAGGCGGAGCGCTCATCGGAGACTTC 660
QY 661 CACCAAGAAATTCGAGAGTCAATTCGGGACACCCCTTGTCTGCGCTCTTTCGAAGACGC 720

DB 661 CACCAAGAAATTCGAGAGTCAATTCGGGACACCCCTTGTCTGCGCTCTTTCGAAGACGC 720
QY 721 GTACAAGATGCTGCAACGATCTTTGCGCCCTCCCGTGGCATGCGCACCGGTTGAAGA 780
DB 721 GTACAAGATGCTGCAACGATCTTTGCGCCCTCCCGTGGCATGCGCACCGGTTGAAGA 780
QY 781 CGCGCGGGAATCGTGAATGCGACTCAGCAGAAAAATCTTTTCACTATCTATACAGCA 840
DB 781 CGCGCGGGAATCGTGAATGCGACTCAGCAGAAAAATCTTTTCACTATCTATACAGCA 840
QY 841 TCACAAGACGAAGAGCTTTTACCCTGCTGACGAACCATCTGCTGACACAGCTGGCGT 900
DB 841 TCACAAGACGAAGAGCTTTTACCCTGCTGACGAACCATCTGCTGACACAGCTGGCGT 900
QY 901 TACGGGCGGCTATGGAAGAGATTTGGCGCTCCGAGGTCAGGAATCTCTGCGCACCG 960
DB 901 TACGGGCGGCTATGGAAGAGATTTGGCGCTCCGAGGTCAGGAATCTCTGCGCACCG 960
QY 961 GAATCCTATGTCGACGCTGACAGCTATATCATGCGGAAGTATATGTAATCTGCGCGG 1020
DB 961 GAATCCTATGTCGACGCTGACAGCTATATCATGCGGAAGTATATGTAATCTGCGCGG 1020
QY 1021 ACGGATCGGCAACCCACAGATTTCTATATCTGCGGCTTCTGTGGCTTACAGTCAAGC 1080
DB 1021 ACGGATCGGCAACCCACAGATTTCTATATCTGCGGCTTCTGTGGCTTACAGTCAAGC 1080
QY 1081 ACGCATGATGCTGTTGATCCCAACCTGGGAGGATACATCAGGAGACAGCGATGAT 1140
DB 1081 ACGCATGATGCTGTTGATCCCAACCTGGGAGGATACATCAGGAGACAGCGATGAT 1140
QY 1141 TTGCAGATAGTCTTGAGGACGAATTCGTCGACAGCGTGCAGCATCTTATGACT 1200
DB 1141 TTGCAGATAGTCTTGAGGACGAATTCGTCGACAGCGTGCAGCATCTTATGACT 1200
QY 1201 GCCATGTTGGAAGAGCGCTTGTCTACATATTTGCCAAGAACCCGACAGATGCACTGGC 1260
DB 1201 GCCATGTTGGAAGAGCGCTTGTCTACATATTTGCCAAGAACCCGACAGATGCACTGGC 1260
QY 1261 CATTCGGTTCCGCGATCCGGAACCCAGGTATACCAACCCATTTACAGAAAGACCCCTG 1320
DB 1261 CATTCGGTTCCGCGATCCGGAACCCAGGTATACCAACCCATTTACAGAAAGACCCCTG 1320
QY 1321 GCACACGAGATTCACCCGAGATGCTTTTCTGATGCGTCCGCTGAGAGTGAGACTC 1380
DB 1321 GCACACGAGATTCACCCGAGATGCTTTTCTGATGCGTCCGCTGAGAGTGAGACTC 1380
QY 1381 TCGTGTCAATCGTCGACCTGCGCTGTTGGCGCAACCCGACCTGGAAGCAAAACACTT 1440
DB 1381 TCGTGTCAATCGTCGACCTGCGCTGTTGGCGCAACCCGACCTGGAAGCAAAACACTT 1440
QY 1441 GGTTCCTCAGAAAGATGTTCAAGAGCGGTAAGTATGCGCGACGCGAGTTCAGATATTC 1500
DB 1441 GGTTCCTCAGAAAGATGTTCAAGAGCGGTAAGTATGCGCGACGCGAGTTCAGATATTC 1500
QY 1501 ACCGACGACTGCTCAAAAGTGAAGCAAGAAATGATGCGGATATGTGGAAGTGGC 1560
DB 1501 ACCGACGACTGCTCAAAAGTGAAGCAAGAAATGATGCGGATATGTGGAAGTGGC 1560
QY 1561 GAGCAACTGGGAGGTTATTTGCCACAGTCCCGGAGTTATGATATCAAGGCTGGC 1620
DB 1561 GAGCAACTGGGAGGTTATTTGCCACAGTCCCGGAGTTATGATATCAAGGCTGGC 1620
QY 1621 ACTTCATCTTGGGGGACTACGCGATTTGCTTCAGCAAGCACTACAGTGGCTGATTA 1680
DB 1621 ACTTCATCTTGGGGGACTACGCGATTTGCTTCAGCAAGCACTACAGTGGCTGATTA 1680
QY 1681 CAATCTGCTGTGTGGGACTTTTGCATCTTATGTTGCAAGCAATGCAACCATCAGAC 1740
DB 1681 CAATCTGCTGTGTGGGACTTTTGCATCTTATGTTGCAAGCAATGCAACCATCAGAC 1740
QY 1741 GGGCTTGGGAGAACCCGACACTTACGTGATGTCACAGCTATCAAGAGGCGGAGAG 1800

Dd	1741	G6GCTTGGCGAGAAACCCGACACTTAAGTCATGTGCCACGCTATCAAGACC	GCAGAGC	1809
OY	1801	CATCATCAATACACTCAAGGGTGGA	CTGACGCAAAAAAATTACAGCGCAGCATCGCAACT	1860
Dd	1801	CATCATCAATACACTCAAGGGTGGA	CTGACGCAAAAAAATTACAGCGCAGCATCGCAACT	1860
OY	1861	TTCGGAAGAAGGCAACAGCACTGTAAACAACCGGCTCAAGTGGTACTTTCAAGTTGAATG		1920
Dd	1861	TTCGGAAGAAGGCAACAGCACTGTAAACAACCGGCTCAAGTGGTACTTTCAAGTTGAATG		1920
OY	1921	CATCTGGTCCCCTACCATGTTGATATGCTACGATAGCGGCTTGAAAGATTTTGTGTATTAC		1980
Dd	1921	CATCTGGTCCCCTACCATGTTGATATGCTACGATAGCGGCTTGAAAGATTTTGTGTATTAC		1980
OY	1981	TGAACCTGTACTTGTCTGTAATAGTTATGACCATATGATTCATGCTTTAAAAA	AAAAAAAAAAAAA	2040
Dd	1981	TGAACCTGTACTTGTCTGTAATAGTTATGACCATATGATTCATGCTTTAAAAA	AAAAAAAAAAAAA	2040
OY	2041	AA	AAAAAAAAAAAAA	2100
Dd	2041	AA	AAAAAAAAAAAAA	2100
OY	2101	AAAAAA	2106	
Dd	2101	AAAAAA	2106	
RESULT 2				
LOCUS	AY291124	2069 bp	mRNA	Linear
DEFINITION	Trimetres ochracea strain MB49 pyranose oxidase (p2o) mRNA, complete cds.			
ACCESSION	AY291124			
VERSION	AY291124.1	GI:31044223		
KEYWORDS				
SOURCE				
ORGANISM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
FEATURES				
source				
gene				
CDS				

BASE COUNT	449 a	659 c	584 g	377 t	NTDSRVBEFKULEPBGCCNIPITAYCANPTLITAMSLAIKSCIEYIKONFTPPSFTBEO
Query Match	10.7%	Score 225.8;	DB 8;	Length 2069;	
Best Local Similarity	51.6%;	Pred. No. 4.3e-36;			
Matches	842;	Conservative	0;	Mismatches 662;	Indels 129; Gaps 9;
QY	268	TGGCTACCAAGAAGAAATGAATGAATTCGAGTAATATATGTAACCCGCTTCGTCAAAGT	327		
DB	306	TGGTGGCCACAAGAAGAAACCGGTGACATACACAGAAGAAACATCGACAAAGTTTGTGAAGT	365		
QY	328	AATCAAGGAGACCTTATCAACAAGTCTGTCTCTGTCTGCAAAACCAGAACTGECCTACACT	387		
DB	366	CATTTCAGGACAACATGATGTGTGTTCGGTTCCCTCATATACCTCGTGTGACACCGCT	425		
QY	388	TGATCCCGGAGCGCTGAGCGCGGCCCGCCCGGAAAGTTACCCATATGGAACGGTAAAAATCC	447		
DB	426	CAGCCCGACGTTCTGTGGCAAGCTTGACG-----TTCCTTCGTCGCAACGGCTCGAACCC	479		
QY	448	TCACCAAGCGGGAATTCGAGAACTTGTCTGCGGAGCCGCTTAACGCGCTGAGAGTCGGCGCAT	507		
DB	480	AGAGAGGAGACCCACTTCTGTAACCTGATGTGTGACGCGGTCAACGCGCGTGTGAGGACAT	539		
QY	508	GAGTACCCACTGTGAGAGCTGCTCCAGCCACGGAATTCATCCACCATTGAAAGTCTCCGGG	567		
DB	540	GTCATCGACATGAGACATCGCCACACCCCGTTTCA-----CCGGGA	581		
QY	568	CATCGCGCGTCCGAAGCTCAGTAACAGACCCGCGCAGAGCAACAAGAGTGGAAAGCACT	627		
DB	582	GCAACGCCCGCTGCTTGTGAAGAGACAGACGCGGACCTCAGACGCTGAGTGGACCGCT	641		
QY	628	TTATTCGAGGAGCGCGCTCTCATCGGAGCTTCCACCAAGAAATTCGAGAGTCAATTGCG	687		
DB	642	CTACACCAAGGCCAGAGTGTACTTCCAACTGTGAGACGAGACCAAGTTCAAGAACTCATCGG	701		
QY	688	GCACACCCCTTGTCTGCGCTCTTGTGCAAGACGCGTATCAAGATGCTCAACTATCTTTCG	747		
DB	702	CCACAAATCTGCTCAACAAGCTCAGCAGAGGAATPAAAGGCCAAGCCGACTTCACAGA	761		
QY	748	CCCTCTCCCGTTGGCATGCACACCGGTTAAGAAAGCGCGCGGAATACGTTCGAATGGCACTG	807		
DB	762	GATTCCACTCGCGGCAACCGCCGG-----AGCCGACCTTCTGTGAATGGAGCTC	812		
QY	808	AGCAAGAAATTTTTTCCACTATCTATCTACAAGATATACAGAGCAAGAAGATCTTTAACCT	867		
DB	813	GCGCAACACTGTTT-----CGACTTCCAGAAACAGGCCGCAACAGCAGCAGC	857		
QY	868	GCTGACGAACATGCTGCACACAGACTGGCGCTTACGCGGCGGGTATGGAAGAAAGATTGG	927		
DB	858	GCGGAGGAGGCGTTCAACGCTCTTCCCGCGGGTCCGTGTGAGCGCGTGTGGCGGACAGC	917		
QY	928	CGCTGCCAGGTCAGGAATCTACTGTGCCACCAAGAAATCTTACTTGTGCACTGTGGACGCTA	987		
DB	918	GTTGAACCTCGAGATCGAGAGCTGTGCACATCCACAGCTCATCTGTGGCGGACCGCTTGA	977		
QY	988	TATCATGCGAGAGATATATGTACTGCGCTGCGGAGCGATTCGAGACCCACAGATTTCTCTA	1047		
DB	978	AATCAAGGCTGACGTACTCTCTCTCACCGCGCGGGCGGTCCACAAACAGCAGCTTCTCTGT	1033		
QY	1048	TAACTCGGGCTT-----CTCTGGGCTACAGGTCAAGCCACAGCAATGACTCGTTGATCC	1101		
DB	1038	GAACTCTGCTTTTGGACAGCTGGGGCGAACCGAACCCCGCAAAACCCAGCGGAGCTCTGCC	1097		
QY	1102	CAACCTGGGGAGGTACATCACAGGAGACCGCATTTTGGCATTTTCCAGATAGTCTTGAAGCA	1167		
DB	1098	GTCCTCTGGGAGCTACATCACCGACGACTGCTGTCTTTCGACAGACCGTGTATGAGCAC	1157		
QY	1162	GGAATTTGGTGCACAGCGT-----	1179		
DB	1158	CGAGCTCATGCAACGCTCAAGTCCGACATACCATCAAGGGGAACCCCTGGCGAGCTCAC	1217		
QY	1180	-----GCGGACAGCATCTTATGGA-CTGCCATGGTGAAGA	1215		

```

Db      1218 GTACAGCGTACGTACACGCCGCGCGTGCAGCAACAGCACCCGAGCTGGTGGAACA 1277
QY      1216 AGCCGTGCTCAATATTGTCAGAGAACCCGACAGATGACATGCGCATTCGGTCCGCGA 1275
Db      1278 GAAGGTGAAMACACATATGACAGACAGAGAGACCCGCTCCCATCCCGTTCCAGGA 1337
QY      1276 TCCGGAACCCAGAGTAACACCCCATTTACAGAAGAACCCCGTGCACACGAGATTCA 1335
Db      1338 CCCCAGCCGACAGGTTCACACTCTGTTCAACCGTCGACCCGTCGACACACTACATCA 1397
QY      1336 CCGCATGCTTTTTCAGGTGCGCGTCCGCTCCTAGGTGAGACTCTCGTGCATCTGCA 1395
Db      1398 CCGCGATGCTTTCAGTTACGGGCGACATGACGAACACATGACTCGCGTCTCATCTGGA 1457
QY      1396 CCGCGATGCTTTCAGTTACGGGCGACATGACGAACACATGACTCGCGTCTCATCTGGA 1455
Db      1458 CTGGCCCTCTTCGCGCGGACGAGCCCAAGAGAGAAACAGCTCGTCTCATCTGGA 1517
QY      1456 TGTTCAGACGGGTACAGTATGCGCGACGCGACGTTACAGATACG--ACCGACACTGC 1512
Db      1518 GATCAGGACGGGTACACATGACATGCGCGACGCGACGTTGACTTCGCGTCCGCGCGCG 1577
QY      1513 GTCAACGTCAGACAGAGAAATGATGCGCGATATGTCGAGAGTGGCCGACACTTGG 1572
Db      1578 CAGGACGAAAGAGGCGGAGACATGATGACCGACATGCTGTTATGTCGCGAAGATCG 1637
QY      1573 AGGTATTTGCCCACGTCGCCCGCGATTTATGATCCAGGCGCTTCATCTCATCTGC 1632
Db      1638 TGGCTTCTTACCGCGCTCCCTCCCGAATTCATGAGCGCTGCTGCTTCTTCCACTTGG 1697
QY      1633 GGGGACTACTCGCATTTGCTTGAC-----AAGCACTACAGTGGCTGATPAACATC 1686
Db      1698 TGGTACGACACGCTGCTGACAGAGAGACACTGCTGCTGCAACAGCACTC 1757
QY      1687 GGTGCTGCGACTTTGCCATCTTATGTTGACGACATGGCACCATCAGACGGGCTT 1746
Db      1758 GCGCGTGTTCGCTTCAAGACCTCTCTCGTGGCTGCGGAAATTCCTACCGCGTA 1817
QY      1747 CCGGAGAACCCGACACTTACGTGATGTCACGCTATCAAGAGCGGAGAGATCAT 1806
Db      1818 CCGCGGAGAACCCGACGCTACCGCATGTCGCTCGGATCAAGAGTGGAGATACATCA 1877
QY      1807 CATACACTCAAG 1819
Db      1878 GCAGACTTCACG 1890

RESULT 3
AF535193      2064 bp      mRNA      linear      PLN 31-DEC-2002
LOCUS      Peniophora sp. SG pyranose oxidase mRNA, complete cds.
DEFINITION      AF535193
ACCESSION      AF535193.1 GI:27436421
VERSION
KEYWORDS
SOURCE      peniophora sp. SG
ORGANISM      Peniophora sp. SG
REFERENCE      1 (bases 1 to 2064)
AUTHORS      Heckmann,D.M., Bastian,S., Rekowski,M. and Giffhorn,F.
TITLE      Pyranose Oxidase of the White-Rot Fungus Peniophora sp. Strain SG:
      Cloning and Characterization of the Gene, Heterologous Expression
      in Escherichia coli, and Properties of the Recombinant Enzyme
JOURNAL      Unpublished
AUTHORS      2 (bases 1 to 2064)
TITLE      Heckmann,D.M. and Giffhorn,F.
JOURNAL      Direct Submission
AUTHORS      Submitted (07-AUG-2002) Applied Microbiology, University of the
      Saarland, Im Stadtwald, 66123 Saarbruecken, Saarland, Germany
FEATURES
source      1..2064
      /organism="Peniophora sp. SG"

```

```

CDS
/mol_type="mRNA"
/strain="SG"
/db_xref="taxon:204723"
26..1897
/EC_number="1.1.3.10"
/note="FAD-oxidoreductase; pox; pyranose: oxygen
2-oxidoreductase; poxSG"
/codon_start=1
/product="pyranose oxidase"
/protein_id="AA013382.1"
/db_xref="GI:27436422"
/translation="MSTSSSDPFNFAKSFPSAAAKRASASLPIRLPDKKVGMD
IKDYIVGSGPIGCTIARELVAGIKVAMFIDIGLSLKGAKTKVETOKIDK
EVNVIQGLQMSVSPVNTLVDTLSPTSMQASTFVRNRSNEPDLRLSGQAVTR
VGGMSTHMTCATPRFDRQRLPLVDDADADAEDMDRLYTKAESYFGCTDOFESIR
HNLVINKLAEEKGORDFOQIPLAATRRSPTEVMSANVTDFLDONRPTDAEERFN
LEPAVACERVRVNAINSEIESIHDLISGRFEELKADYVLTAGAVHNTOLVNSGF
GOLGRNPTNPPELLPLSGSYITEOSLYFCQVYMETELIDSKDPTIRGTGELTYS
VYVTPGASTNKHHPDMWNEKVKHMHMOHODPLIPPEDEPOVTLTFQSPHWHQIH
RDAFSYGAQVQSIDSNLIVDMRFQGTERTKEENKLMFSDKITDAYNMQPTDFRPA
GRTSKAEEDMMDKCVMSAKIGFLPGLSPQFMEBGLVHLGTRHMGFDEKEDNCV
NDSRVFGRKNLFLGCGNIPYAVGANPTLFRMSLAIKSCEYIKQNFPTSPPTSAO"
poly_site      2047
BASE COUNT      462 a      652 c      579 g      371 t
ORIGIN
Query Match      10.5%; Score 221; DB 8; Length 2064;
Best local Similarity 51.4%; Pred. No. 4.2e-35;
Matches 839; Conservative 0; Mismatches 665; Indels 129; Gaps 9;
QY      268 TGGCTACACAGAGAAATGAAATGAGTTCACAGAAAGATATTGACCGCTTGCAATGT 327
Db      283 TGGTGCACAGAGAAAGAACCCGTCGATGACAGAAAGAACATGTTGTACACT 342
QY      328 AATTCAGAGCGCTTACACAACAAGTCTCTTCCTGAGAAACGAAAGTGCCTACACT 387
Db      343 CATTCAGGACAAATTGATCTGTTTCCTTCCTGATACCTTCGTGTCGACACGCT 402
QY      388 TGATCCCGAGCGCTGAGGCGCCCTTGAAAGTTTCAGCATATGAAAGGTAAAAATCC 447
Db      403 CAGCCGACGCTGTGGCAAGCTTCGACG-----TTCCTGTCGCGCAACGCGCTCAACCC 456
QY      448 TCACACGCGGGAATTCGAGAACTTGCTCTCGGAGCCGCTAACCGCTGGAGTGGCGCAT 507
Db      457 AGAGCAGGACCCGCTTCGTACCTCAGTGCAGCGGTCACGCGCTGTGGAGGCAAT 516
QY      508 GAGTACCCACTGAGAGCTGCTCCACGCGACGAGATTCATCACCATGAGAAAGTCTCCGCG 567
Db      517 GTCTAGCACTGAGATGCGCCACACCCCGTTTCA-----CCGCA 558
QY      568 CATCGGCGCTCGAAGCTCAGTAACGACCCGCGAGAGACGAAGAAGAGTGGAGAGCT 627
Db      559 GCAAGCGCCGCTGCTTGTGAAGACGAGCGGACGCTGAGTGGAGCGGCT 618
QY      628 TTATTCGAGGCGCGAGCGCTCTCATCGGACTTCACCAAGAAATTCGAGCATTCATTCG 687
Db      619 CTACACCAAGGCGGAGCTTACTCTCAGACTGGAGCGCCAGTTTCAGAGACTCATATCG 678
QY      688 GCACACCCCTGTTTCGCGCTCTTTCACAGACCGTACAGATGATGCTCAAGCTATCTTCG 747
Db      679 CCACACCTCTGCTCTCAACAGCTCGCGAGGAATACAGAGCCCAACGCACTTCACAGA 738
QY      748 CCGCTCCCGTTGGCATGCGCACCGGTTGAAGACCGCGGAATACGTCGATTCGACTC 807
Db      739 GATTCACACTCGGCGGACCGCGG-----AGCCGACCTTCGTGATGAGCTC 789
QY      808 AGCAGAAATCTTTTCCACTCTATCTACACGATACAGACGAAGAAAGCTTTTACCTT 867
Db      790 GCGCAACTCTGTTT-----CGACCTTCAGAACAGGCGCAACGAGGACG 834
QY      868 GGTACGAAACATCTGCTGACACGACTGCGCTTACGGCGGGGTATGAGAAAGATTGG 927
Db      835 GCGGAGAGCGCTTCAACTCTTCCCGCGGTGCTGTGAGCGGCTGTCGCAACGC 894

```

OY	928	CGCTCCGAGGCTACAGAAATCTACAGCGCCACCACGAAATCCTGATTGTGAGAGTGCACAGCTA	987
OY	928	CGCTCCGAGGCTACAGAAATCTACAGCGCCACCACGAAATCCTGATTGTGAGAGTGCACAGCTA	987
Db	895	GTTGAACTCGGAGATCGAGAGCTCTGCACATCCAGACCTCATCTCTGGCGCACCGCTTTGA	954
OY	988	TATCATGCGAAGGATATATGACTGCGCTCGGGAGCCGATCGGCAACCCACAGATCTCTA	1047
Db	955	AATCAAGGCTGACGATGATGCTGCTCACCGCGGGCGGTTCCAAACACGACGCTTCTGT	1014
OY	1048	TAACTCGGGCTT-----CTCTGGGCTACAGGTACAGCCACGCAATGACTGTTGATCC	1101
Db	1015	GAACCTCTGGCTTTGGACAGCTGGGGCCGACCCCAACCCCAAAACCCACCGGAGCTGCTGC	1074
OY	1102	CAACCTGGGGAGGTACATCACGGAGACCCGATGGCATTTTGGCAGATAGTCTTGAGGCA	1166
Db	1075	GTCCTCGGGGAGCTACATCACCGAACAAGTCTGCTCTTCTGCGACGAGCTGATGAGCAC	1134
OY	1162	GGAATTCGTGCACACGT-----	1179
Db	1135	CGAGCTATCGACAGCGTCAAGTCCGACATGCATCAAGGGGAACCCCTGGCGAGCTGAC	1194
OY	1180	-----GCGGACGATCTTATGGA-CTGCCATGGTGAAGA	1215
Db	1195	GTACAGCGTGAAGTACACGCCGGGGCGCTGCACCAACGAACCGGAGCTGGTGAAGA	1254
OY	1216	AGCGCTGCTCAACATATTGGCAAGAACCCGACAGATGACATGCCCCATCTGCTTCCGGGA	1275
Db	1235	GAAGGTGAAAAACCATGATGACAGCACCGAAGAGACCCGCTCCCGATCCGCTTCGAGGA	1314
OY	1276	TCCGGAACCCAGGTAAACMACCCCATTTACAGAAGAACACCCCTGGGACACGACGATTTCA	1335
Db	1315	CCCCGAGCCGAGGTATACACTCTGTTCCAAACCGTCCGACCCGTTGGGACACTCAGATCCA	1374
OY	1336	CCGGGATGCTTTTTCGATACGCTGCCCTGCTCGATGAGTGCATCTGTTGATCTGCA	1395
Db	1375	CCGGGATGCTTTTTCGATACGCTGCCGAGTGCACAAAGCATTCGACCTGCTCATCTGTGA	1434
OY	1396	CCTGGCGGTTTGGCGGACGACCCCGAAGAACAAACCTTTTGGTTTTCGAGAAAGA	1455
Db	1435	CTGGCGCTTCTTCCGCGCGAGGAGCCCAAGAGAGAGAACAGCTCTGTTCTCGACAA	1494
OY	1456	TGTTCAAGACGGGTACAGTATGCCGACCGACGCTTCAGATATCG--ACCAGACATGC	1512
Db	1495	GATCAGGAGCGGTACACATGTCGCGACGCCACGTTTCAGCTTCCCGGGCGGGCG	1554
OY	1513	GTCAAAGCTGAGAGCAAGGAAATGATGGCCGATATGTGCCAAGTGGCGAGCAACTTGGG	1572
Db	1555	CACGAGCAAGGAGCGGAGACATGATACCGACATGTGCTTATGTGCGGGAAGATCGG	1614
OY	1573	AGGTTAATTGGCGACGTCGCCCGCACTTATGATGCACGAGCCCTGACATCTCATCTGC	1632
Db	1615	TGGCTCTACCCCGGCTCTCTCCCGCAATTCATGAGCGCTGTCTTGTCTTCACTCTGG	1674
OY	1633	GGGGACTACTCGATTGGCTTCGAC-----AAGGCAACTACAGTGGCTGTAAACACTC	1686
Db	1675	TGTTACGACCGCATGGGCTTCGACAGAAAGAGGAACTAGTGGCTGTCACACGAGACTC	1734
OY	1687	GCTGGTCTGGGACTTTGGCATCTTTATGTTGACAGGCAATGGCACATCAGAGGGGCTT	1746
Db	1735	GCGGCTGTTGGCTTCAAGAACCTTCTCTCGGTGGCTGCGGAAACAATTCCTACCGGCTA	1794
OY	1747	CGGCGAGAACCCGACACTTACGTGATGTGCACGCTTCAAGAGCGGAGAGACATAT	1806
Db	1795	CGGGCGAACCCGACGCTCACCGGAATGTGCTGCGCATCAAGAGTTGCGGATACATCA	1854
OY	1807	CAATPACACTCAAG 1819	
Db	1855	GCAGAACTTCACG 1867	

DEFINITION	Sequence 1 from patent US 6146865.
ACCESSION	ARI41573
VERSION	ARI41573.1
KEYWORDS	GI:15101089
SOURCE	Unknown.
ORGANISM	Unknown.
REFERENCE	Unclassified.
AUTHORS	1 (bases 1 to 1869)
TITLE	Christensen,S., Lassen,S.,Plensted, and Schneider,P. Nucleic acids encoding polypeptides having pyranose oxidase activity
JOURNAL	Patent: US 6146865-A 1 14-NOV-2000;
FEATURES	Location/Qualifiers
source	1..1869
BASE COUNT	/organism="unknown"
ORIGIN	397 a 622 c 548 g 302 t
Query Match	10.1%; Score 212.4; DB 6; Length 1869;
Best Local Similarity	51.5%; Pred. No. 2.4e-33;
Matches	841; Conservative 0; Mismatches 656; Indels 135; Gaps 10.
QY	269 GGCTACCAAGAAAGATGAATGAGTTCTCCAGAAAGATATTGACCGCTTGCTCAATGTA 328
DB	259 GGCTACCAAGAAAGAAACACCGCTGAGTACCAAGAAACATCGACAAATTTGTAATGTT 318
QY	329 ATCAAGGAGGCGTTACACACAGTCTCTTCCCTGTCACAAACACAGACGTCGTCACTT 388
DB	319 ATACAGGGGCACTTATTATCCCGCTCTCGGTGCCCGTCAACACAGATGTGTTGACACGCTA 378
QY	389 GATCCCGGAGCGCTGAGCGCGCCCGCTGGAAGTTTCAGCCATATCGAACGCTAAATCT 448
DB	379 AGCCCGGCGTATGGCAAGCTTCGACG-----TTCTTCGTCGCCCAAGGGGCAATCCA 432
QY	449 CACCAAGCGGGAAATTCGAAACTTGTCTCTCGGAGGCGCCGTAAAGCGCTGAGAGTGGCGCATG 508
DB	433 GAGCAAGACCCGCTGCGCAACCTTAAGTGGCCAGGCGGTCACCCCGCGTGTGGCGGCATG 492
QY	509 AGTACCCACTGGAGCGTGTCTCAGCGCAAGATTCATCCACCATTGGAAGTCTCCGGGC 568
DB	493 TCTACGCACTGGAGCGTGGCGACGCGCGGCTTCGAAAGCTGCGA----- 537
QY	569 ATCGGCGCTCCGAAGCTGATGACAGACCCCGCAGAGACGACGACAAAGACTGGACAGCTT 628
DB	538 ---GCGCGCTGCTGTGTAAGAGACACTCCAAAGGGGCGAGCGCGGATGGGACAGGCTC 594
QY	629 TATTCCGAGGCGCGAGCGCTGTCATGCGGGACTTCACACCAAGAAATTCGAGAGCTATGG 688
DB	595 TACAAGAAAGGCGCGAGTCTACTTCAAGACCGGACAGACCCAGTTGCGCGAGTCAATCCGC 654
QY	689 CACAACCTTGTCTGCGCTCTTTGCAAGACGCGTAAAGAGATCGTCAACGTATCTTTGCG 748
DB	655 CACAACCTCTGTGCTCAAGAACTGTGAGGAGGAGTAAAG---GCGTTCGGAGCTTCCAG 711
QY	749 CCTTCCCGGTGGGATGCGACCGCGGTGAAGAACCGCGGGAATTCGTGATGATGACACGTA 808
DB	712 CAGATCCCGCTCGGGGCGAGCGGCGAG-----GCCGAGCTTGTGAGTGGAGGCTCG 765
QY	809 GCAGAAATCTTTTTCACACTTATCTACACGATGACAAAGCAAGAAAGCTTTTAACTCTG 868
DB	766 GCGCAACACGCGTT-----CGATCTCGGAACCGGCGCAACAAAGACGCG 810
QY	869 CTGACGAACCATCCGCTGACACGACTGGCGCTTACGGCGGGGTATGAGAAAGAAATTGGC 928
DB	811 CCGAAGCAGCGCTTCAACCTCTTCCCGCGCTCCGCTGCAAGAACGTGAGGCGGATTAAC 870
QY	929 GCTCCGAGGTCAGGAATCTACTGGCCACCGAAGATCTTAAGTGGACACTGGACATAT 988
DB	871 GCGAATCTCGGAGATCTGAGCGCTGATGTCCGCACTCCACGGGGGCGAGAGCATCAC 930
QY	989 ATCATGGCGAAGTATATGTAAGTGGCGTGGGAGCGATCGGCAACCCACAGATTTCTTAT 1048
DB	931 ATCAAGGCGAAGGTATATCTCTACCGCGGCGCGGCTGTCACAAAGCGGACCTCTCGCG 990

QY 1049 AACTGGGGCTT-----CTTGGGCTACAGGTACAGCCAGCAATGAC--TCGTTGATC 1099
 Db 991 GCCTTGATGTCGGGAGCTGGGTCTGCCGAGCCCGCAAGCGCTGCCGTCTGCTG 1050
 QY 1100 CCCAAGCTGGGAGGTACATACACGAGCAGCCGATGGCATTTTGGCAGATGCTTGAGG 1159
 Db 1051 CCGTACCTGGGAGCCACATCACAGAGACGCTGTCTTGCAGACCGTCATGAGC 1110
 QY 1160 CAGATATGTCGACAGCGTGGCGCAGATC----- 1190
 Db 1111 AGGAGCTCATCAAGGTGTCAACCGGGATATGACATTTGTGGCAAGCCCGCCACCG 1170
 QY 1191 -----CTTANGACTGCCATGTTGAA 1213
 Db 1171 GACTATAGCTACGATATACCCCGGCAACCCGAAACAGACAGCAGCCGAGCTGTGGAAC 1230
 QY 1214 GAAGCGGTGTCATCATATTTGCAAGAACCCGACAGATGCACTGCCCATTCGTTCCGC 1273
 Db 1231 GAGAAAGTGAAGAGACATGATGACACACAGAGAGCCCGCTCCGATCCGTTGAG 1290
 QY 1274 GATCCGGAACCCAGGTAAACAACCCATTTACAGAAACACCCCTGGACACGAGATT 1333
 Db 1291 GACCTGAGCCGAGGTACACAGCTGTTTCAAGGCAACGACCCATGGCAGACCCGAT 1350
 QY 1334 CACCGGATGCTTTTTCGACGGTCCGCTCGCTGAGGTGACCTCTGTCATGCTC 1393
 Db 1351 CACCGGAGCGCTTACGATACGGCGCCGTGACAGAGATGATGATCGCGGCTCATGCTC 1410
 QY 1394 GACCTGCGTGGTTGGGCAACCCGACCCCTGAAGCAAAACCTTTTGGTTTCCGAAAC 1453
 Db 1411 GACTGGCGTCTTTCGAGCAGCACCGAGCCAGAGAGAGAAAGCTATGTTGTTGAGC 1470
 QY 1454 GATGTTCAAGAGCGGTACAGTATGCCGACCGACCTTCAGATATTCGACCAAGCTGCG 1513
 Db 1471 AAGATCAGGAGCGGTACAAACCTCCGAGCCGAGCTTGACTTCGCTCCCGGGG-- 1528
 QY 1514 TCAAAGTGAAGCAAGAAATGATGCCGATATGTCGAACTGGCGAGCAACTTGGGA 1573
 Db 1529 ----GCCGGAAGCGAGAGACATGATGACCGCATGTGCTATGTGCGGAGAAATCGGT 1584
 QY 1574 GGTATTTTCCCAACGCCCCCGGAGTTATGATGATCCAGGCTTCACATTCATCTGCGG 1633
 Db 1585 GATTTCTCTGCTGGGTCTTACCCACAGTTTCATGAGCCCGGCTTCTCTGCACTTGGT 1644
 QY 1634 GGGACTACTCGCATGTTGGCTTGACAA-----GGCAACTACAGTGGCTGATAACAAC 1687
 Db 1645 GGGAGCGACCGCATGGGCTTGACAGAGGGGAGCAAGTGTGCTGACACCGATCA 1704
 QY 1688 CTGCTCTGGGACTTTGCCATCTTTATGTTGACAGCAATGCAACATCAGAGCGGCTTC 1747
 Db 1705 CCGCTCTTGGGCTTCAGAAACCTTCTCTGCGGCGCTGCGGAAATCCCAACCGGTAC 1764
 QY 1748 GGGGGAACCGCACTTACGTGATGATGTCGACACGCTATCAAGAGCCGAGAGCATATC 1807
 Db 1765 GCGGGAACCGCACTTACCGCAATGATGCTTGGCATGAAGAGTGCAGATCAAG 1824
 QY 1808 AATACACTCAAG 1819
 Db 1825 AAGAACTTCGAG 1836
 RESULT 5
 LOCUS E11766 1869 bp RNA linear PAT 29-SEP-1997
 DEFINITION cdna of pyranose oxidase.
 ACCESSION E11766
 VERSION E11766.1 GI:22025402
 KEYWORDS JP 1996205861-A/1.
 SOURCE unidentified
 ORGANISM unidentified
 REFERENCE 1 (bases 1 to 1869)

AUTHORS Nishimura, I., Okada, K., Nanbara, T., Kawai, G., Koyama, T. and Suzuki, M.
 TITLE NEW PYRANOSE OXIDASE, PYRANOSE OXIDASE GENE, NEW RECOMBINANT DNA AND PRODUCTION OF PYRANOSE OXIDASE
 JOURNAL Patent: JP 1996205861-A 1 13-AUG-1996;
 KIKKOMAN CORP
 COMMENT OS Coriolus versicolor
 PN JP 1996205861-A/1
 PD 13-AUG-1996
 PE 24-MAY-1995 JP 1995124835
 PR 07-DEC-1994 JP 94P 304086
 PI NISHIMURA IKUKO, OKADA KIMIHARU, NANBARA TOMOYUKI, PI KAWAI GENSHIRO,
 KOYAMA TAJIJI, SUZUKI MASARU
 PC C12N9/04, C12N15/09, (C12N9/04, C12R1:19), (C12N15/09, C12R1:01);
 CC strandedness: Double;
 CC topology: Linear;
 FH key Location/Qualifiers
 FT source 1..1869
 FT /strain='ps4a', /organism='Coriolus versicolor' FT
 FEATURES
 source 1..1869
 /organism='unidentified'
 /mol_type='genomic RNA'
 /db_xref='taxon:32644'
 BASE COUNT 408 a 593 c 529 g 339 t
 ORIGIN
 Query Match 10.0%; Score 210.2; DB 6; Length 1869;
 Best Local Similarity 50.4%; Pred. No. 6, 9e-33;
 Matches 857; Conservative 0; Mismatches 713; Indels 129; Gaps 9;
 QY 200 GAGATGGAGCGCTGATAGCTTCTACGCTGTTAATGCCGAAGAGAACTGACGTCC 259
 Db 190 GACCTGTGGAAGCCCGTTACAAAGTCCCATGTTGACATCGGGAAATTTGACTTGGC 249
 QY 260 TACGTTCTGCTGCTACCAAGAAATGAATGAAATGCAATTCAGAAATTTGACCGCTTC 319
 Db 250 CTGAAGATCGTGTCCACAAAGAAACACCGCTGATACCAAGAAACATTTGACAGTTT 309
 QY 320 GTCAATGTAATCAAGGAGCGCTTACAACAAGTCTGTTCTGTCAGAAACGAAGCTG 379
 Db 310 GTGAAGCTATCAGAGGCGCAATGATGTGTTCCGTTCCGTCATATACCTCGTGAATC 369
 QY 380 CCTACACTGATGCCGAGCGCTGGAGCGGCGCCCTCGAAGTTACCCATATGAAAGGT 439
 Db 370 GACAGCTCAGCCCGACGCTTGGCAAGCTTCAATG-----TTCTTCTCGCAATGCG 423
 QY 440 AAAAATCTTACCAGCGGGAATTCGAAACTTGTCTGCGAGGCGCTTAACGCGTGAATC 499
 Db 424 TCGAACCAGAGAGAGACCGCGTTCGTAACTAGTGTGACAGCGGTGACCGCTGTGTC 483
 QY 500 GCGGCTAGTACACCACTGAGCTGCTCAGCGCAGGATTCATCCACCATGGAAGT 559
 Db 484 GGAGCATGTCCACGACATGACATCGGACACCCGCTTTGA----- 527
 QY 560 CTCGCGGATGAGCGCTCCGAAGCTCAATACGACCGGCGAGAGGAGCAAAAGAGTGG 619
 Db 528 --CCGCGAGACAGCGCGCTTGTCTGTAAGAGAGACCAAGAGCGCTGACGCGCGATGG 585
 QY 620 AAGAGCTTTATTTCCAGAGCGCGCTCTCATCGGACTTCCACCAAGAAATTCAGAGAG 679
 Db 586 GACCGCTGTACACCAAGGCGCGATATCTTCAAGACCGGAGCGGACAGTTTCAAGAG 645
 QY 680 TCAATTCGACACACCTTGTCTGCGCTCTTTGCAAGACGCGTCAAGAGATGTCACGT 739
 Db 646 TCGATCCGCCACAACCTGTGCTCAACAAGCTGCGGAGGAATACAAAGTGCAGCGGAC 705
 QY 740 ATCTTTGCCCTCTCCGTTGGCATGACCAAGCGGTTGAAGAAGCGCGGGAATATCGTGA 799
 Db 706 TTCCAGCAGATCCCGCTCGGCGCAACGCTCG-----CAGTCCGACCTTCGTGAG 756

Db 215 GAGCTGTCGAGAGCCGGTTACAAAGTCCGATGTTGCATCGGGGAATTTGACTTGGC 274
 QY 260 TACGTTCTGGCTACCAAGAGAAATGAATGAGTTTCCAGAAAGATTTAGACCGCTTC 319
 Db 275 CTGAAGATCGGTGCCCCCAAGAAAGACACCGTGCATACCAAGAAACATTGACAACTTT 334
 QY 320 GTCAATGTATCAAGAGAGGAGCTTACAAACAATCTCTGTCTCTGTACAGAAACAGAACGTG 379
 Db 335 GTGAAGTCATTCAGGGCCAAATTGATGTCTGTTCCGTTCCCGTCAATACCTCTGTATC 394
 QY 380 CTACACTTATATCCCGAGGCTTGAGCGGCCCCCTGGAAGTTTCAGCCATATGCAAGCT 439
 Db 395 GACACGCTGACGCCGACGCTTGTGCAAGCTTTCATCG-----TTCCTTGTCTCCGAATGGC 448
 QY 440 AAAATTCCTCACCAGCGGGAATTCGAGACTTGTCTCGAGAGCGCGTAAAGCGTGGATC 499
 Db 449 TCGAACCAGAGAGAGACCGCGCTTCGTAACCTCAGTGTGTAAGCGGTGACGCGTGTGTC 508
 QY 500 GGGGGAGTAGTACCCACGTGAGCTGTCCACGCCAGATTCATCCACCCATGGAAGT 559
 Db 509 GAGGAGATGTCCAGGACCTGAGATGCGGACACCGGCTTGA----- 552
 QY 560 CTCGCGGAGATCGGCGCTCGAAGCTAGTAAGACCCGCGAGAGACAGCAAGAGTGG 619
 Db 553 --CCGGAGAGAGCGCCGCTGTCTGTGAAGAGACGACGAGCGCTGACGAGCGCGAGTGG 610
 QY 620 AACGAGCTTATTCGAGAGCGGAGCGCTCATCGGAGCTTCCACCAAGAAATTCGAGAG 679
 Db 611 GACCGGCTGTACACCAAGGCGCGAGTCACTTCAAGCCGAGGAGCCAGTTCAGAGAG 670
 QY 680 TCAATTCGAGACACCTTGTCTGCGCTCTTGTGCAAGACGCTACCAAGATTCGTAAGT 739
 Db 671 TCGATCGGACCAACCTCGTGTCAACAGACTCGCGAGAGATTAAGAGTACAGCGGAG 730
 QY 740 ATCTTTCGCGCTCTCCGCTTGCGATGCGACCGGTTGAAGAGCGCGGAGTAATGAGTGA 799
 Db 731 TTCCAGAGATCCCGCTCGCGGCAACGCTG-----CACTCCGACCTTCTGTCGAG 781
 QY 800 TGGCACTCAGCAGAAATTTTTCACACTATCTACAAAGTATGCAAGAGAAAGAACTC 859
 Db 782 TGGAGCTCGGCGAAGACCGTGT-----CGACCTCCAGAAACAGCGCGAAGC 826
 QY 860 TTTACCTGCTGAGCAACATCGCTGCACAGACTGGCGCTTACGGGCGGGTATGAGAG 919
 Db 827 ACGGAGCGCGGATGAGCGCTTCAACCTCTCCCGGTTGATGTGAGCGCGTGTG 886
 QY 920 AATATTGGCGCTCGCGAGGTGACAGAACTCTACTGGCCACAGAAATCTTACTTGCACGTG 979
 Db 887 CGCAACACGTGCAACTCCGAGATCGAGAGTCTGCACATCCACGACCTCATCTCGGCGAC 946
 QY 980 GACAGCTATATCAATGCGAAGATATATGTAATGCTGGCGTGGGAGGAGTCCGCAACCCACAG 1039
 Db 947 CGCTTGAATCAAGAGACAGCGTGTCTTACAGCCCGGCGGTCTCCACAGCGCGAG 1006
 QY 1040 ATTCTCTATATCTGGGCTT-----CTCTGGGCTACAGGTCAAGCCAGCAATGACTCG 1093
 Db 1007 CTCTCGTGAATCTGCTTGGACAGCTGGCGCGCGGAGCCCGGAGACCCGCGCGAG 1066
 QY 1094 TTGATCCCAACTCGGGAGGTATACGAGAGACGCGATGTCATTTTGGCAGATATGTC 1153
 Db 1067 TTCTCGCGCTCTGGGAACCTACATCACGAGAGTCTGCTCTTCTCGAGAACGCTG 1126
 QY 1154 TTGAGGAGGAATTCGTGACAGCGTG-----CGCGAGATTCCTTATGAGCTGGC----- 1203
 Db 1127 ATGAGCACCGAGCTCATCGACAGGTCAGTCCGACATGATCATAGGGCAACCCCTGGC 1186
 QY 1204 -----ATGG 1207
 Db 1187 GATCTGGGGTACAGCGTACAGCCCGGGCGGAGACCAAGCAACCGGAGATGG 1246
 QY 1208 TGGAAAGAGCGCTTGTCTCAATATTTGCCAAGAACCCGACAGATGACCTGCCATTCG 1267
 Db 1247 TGGAGGAAAGGTGAAGAACACATGATGACGACACGAGAGACCGCTTCCAAATCCG 1306

QY 1268 TTCCGCAATCGGAAGCCCAAGATTAACACCCATTTACAGAGAACACCCCTGGCACAG 1327
 Db 1307 TTGAGAGACCCCGAGCCGAGAGTACACACCTTGTTCAGGACATGCGACCCGTGGACACT 1386
 QY 1328 CAGATTCACCGGATGCTTTTTCGTACGAGTCCGCTGCTGATGAGTGTGATCTGCTGTC 1387
 Db 1367 CAGATTCACCGGATGCTTTTTCGTACGAGTCCGCTGCTGATGAGTGTGATCTGCTGTC 1426
 QY 1388 ATGCTGACCTGCGCTGCTTGTGCGCAACCGACCTTAAGCAAAACCTTTTGGTTTC 1447
 Db 1427 ATCTGACATCGGCGCTTCTTCGCGGAGGAGGCAAGAGAGCAAGCAAGCTCTGCTTC 1486
 QY 1448 CAGAGATGATTAAGAGGAGTACAGATGAGTCCGAGCGGAGCTTCAATATGACAGCAGC 1507
 Db 1487 TCGGACAAATTTACGACAGTACATGCGCGAGCGGAGCTTCCGCTTCGCGG 1546
 QY 1508 ACTG--CGTCAACGTAAGAGCAAGAAATGATGGCCGATATGTGCGAAGTGGCAGC 1564
 Db 1547 GCGGCGCGACAGCAAGAGGAGGAGACATGATGACCGATATGTGCTTATGTCGCG 1606
 QY 1565 AACTTGGAGATTTTGGCCAGCTCCCCCGAGTTTATGATTCAGGCTTGCCTT 1624
 Db 1607 AAGATTTGTGTGCTTCTGCGCGGCTCCCTCCGCAATTCATGAGGCGCGCTTGTCTT 1666
 QY 1625 CATCTTCGCGGACTACTTCGCAATGCTTGCAC-----AAGGCACTACAGTGGCTGAT 1678
 Db 1667 CACTCGGTGTGTACGACACCGCATGGCTTGCAGCAGAGAGCAAGTGTGCTGAC 1726
 QY 1679 AACAACTGCTGTCTGTGGGACTTTCGCAATCTTTATGTTGTGAGCAATGAGCAATCAG 1738
 Db 1727 ACGGACTGCGGCGGCTTGTGCTCAAGAACTGTTCTCGGTGCTGCGGAAACTTCCC 1786
 QY 1739 ACGGCTTGGCGGAGAACCCGACACTTACGTGATGTGCCAGCTATCAAGAGCGGAG 1798
 Db 1787 ACGCGTACGCGGAGAACCCGACGCTACCGCAATGTGCTGCGATCAAGAGTGGAG 1846
 QY 1799 AGCATCATCAATACATCA 1817
 Db 1847 TACATCAAGAACTTCA 1865

RESULT 8
 E32737
 LOCUS E32737 1902 bp DNA linear PAT 18-JUN-2001
 DEFINITION Bannus cornucopiae-derived antitumor protein and gene thereof.
 ACCESSION E32737
 VERSION E32737.1 GI:13017467
 KEYWORDS JP 1999J15096-A/1.
 SOURCE Pleurotus cornucopiae (cornucopia mushroom)
 ORGANISM
 Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
 Agaricales; Pleurotaceae; Pleurotus.
 REFERENCE
 1 (bases 1 to 1902)
 Akihiro, M., Koji, I., Kazuaki, K., Tomohide, S. and Ryoko, A.
 Panus cornucopiae-derived antitumor protein and gene thereof
 Patent: JP 1999J15096-A 1 16-NOV-1999;
 NYUFUD KURIKISHON GIUTSUKEKINYUMINAI
 COMMENT
 OS Pleurotus cornucopiae
 PN JP 1999J15096-A/1
 PD 16-NOV-1999
 PF 07-AUG-1998 JP 1998236349
 PR
 PI AKIHIRO MORITA, KOJI IZUMO, KAZUAKI KIDO, TOMOHIDE SAKA, PI
 RYOKO AOKI
 PC C07K14/375, A61K35/84, A61K38/00, C07K16/14
 CC
 FH Key Location/Qualifiers
 FT source 1..1902
 FT Location/Qualifiers
 1..1902
 /organism="Pleurotus cornucopiae"
 /organism="Pleurotus cornucopiae"

/mol_type="genomic DNA"
/db_xref="taxon:5321"
BASE COUNT 437 a 556 c 479 g 430 t
ORIGIN

Query Match 8.3%; Score 175.6; DB 6; Length 1902;

Best Local Similarity 57.1%; Pred. No. 8.8e-26; Mismatches 264; Indels 9; Gaps 2;

Matches 363; Conservative 0; Mismatches 264; Indels 9; Gaps 2;

1205 TGGTGAAGAAGCCGTTGCTCAACATATTTGCCAAGAACCCGACAGTACCTGCCATT 1264
1249 TGGTGAAGAAGCCGTTGCTCAACATATTTGCCAAGAACCCGACAGTACCTGCCATT 1308
1265 CCGTTCGCGGATCCGGAACCCGAGTAACACCCATTTTACAGAAAGAACCCCTGGCAC 1324
1309 CCGGTCGATGACCCGAGCCAGCAAGTACAGACATCTTCCAGACACATCATATGCGAC 1368
1325 AGCAGATTCACCGGATGCTTTTTCGTAACGCTGCGCTGCTCGAGTGCATCTGCT 1384
1369 ACTCAGATTCACCGGATGCTTTTTCGTAACGCTGCGCTGCTCGAGTGCATCTGCT 1428
1385 GTCATCGTGAACCTGCTGCTTTGGCGCAACCGACCTGAGCAAAACCTTTGGTT 1444
1429 CTTGTTGTTGACTGGCGCTCTTGGTGGCACTGAACCTGTGAGAGAAAGACCTGTGG 1488
1445 TTCCGAAGATGTTCAAGACGGGTACAGTATCCGCGACCGAGCTTCAGATATCG--A 1501
1489 TTCTCAAAACAGATCACTGATGATCAACCTGCCACAGCCCATTCAGCTTCGGGTTC 1548
1502 CCCAGCAGCTGGTCAAGGTGAGAGCAAGAAATGATGCGCATATGCGCAATGGGCG 1561
1549 CCCAGGCGCGACAGCCGAGGCGGAGCTGATGATGCGCATATGCGCAATGGGCG 1608
1562 AGCAACTTGGGAGTTATTTGCCAGTCCCGCCGAGTTATGATTCAGGCGCTTGA 1621
1609 ACCAAGTGGTGGCTTCTGCGGTTGATTCGCGAGTTATGATTCGCGCTTGA 1668
1622 CTTGATCTTGGGAGTACTGCTGCTTGGCTTGCAGAAAGCACTACAGTGC-----T 1675
1669 CTCCACCTTGGTGAACCCGAGCTGAGCTTGTGATGAGGAAGCTGATAGGCGGTGTC 1728
1676 GATCAACACTGCTGGTGGGAGCTTTGGCAATCTTTATGAGGAGCAATGCCACATC 1735
1729 GACACCACTCAAGGCTCTTGGCATGAGAACTGTTCTTGGGCTTGGCGGCAACAT 1788
1736 AGGACGGGCTTGGGAGAACCCGACACTTACGTGATGTGCGACGTATCAAGACGCG 1795
1789 GGCACGGGTAGCTCGAAGCCGACGCTCACTGGGCTTGGCTGCAATCGAGAGCTGT 1848
1796 AGGACATCATATACACTCAAGGCTGAGTGC 1831
1849 AAGTATATCAGGAATATTTTCAGGCCAAGTCTGTC 1884

RESULT 9
LOCUS ARI170485 1701 bp DNA linear PAT 17-DEC-2001
DEFINITION Sequence 2 from patent US 6291648.
ACCESSION ARI170485
VERSION ARI170485.1 GI:17908444
KEYWORDS
SOURCE .
ORGANISM Unknown.
REFERENCE Unclassified.
AUTHORS 1 (bases 1 to 1701)
TITLE Kawamura,Y., Morita,A., Izumo,K. and Saka,T.
JOURNAL Antitumor protein and corresponding gene sequence isolated from
FEATURES matsutake mushrooms
PATENT: US 6291648-A 2 18-SEP-2001;
LOCATION/Qualifiers
1..1701
BASE COUNT 411 a 463 c 432 g 395 t

ORIGIN

Query Match 7.2%; Score 152.6; DB 6; Length 1701;
Best Local Similarity 53.5%; Pred. No. 4.6e-21;
Matches 394; Conservative 0; Mismatches 304; Indels 39; Gaps 2;

1106 CTGGGAGGTACATACGAGCAGCCGATGTCATTTTGGCAGATATGCTTGAAGCAAGAA 1165
964 CTTGGTCCGTACCTCAGGCAACGATGATGATTTTTCGATGATGCTTCTCAAGAGGGCC 1023
1166 TTGCTGACAGGCTGGGAGCATGCTTATGACATGCGCATGCTGGAAGAAAGACCTTGGT 1225
1024 ATATGATGATGCTGCTACTGATCCT-----CGCTTGGCTGCAAGGTTGAG 1071
1226 CAACATATTTGCCAAGAACCCGACAGATGATGATGATGATGATGATGATGATGATGAT 1285
1072 GCGCAAGAAAGAACACCCGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1131
1286 CAGGTACACACCCATTTTACAGAGAACACCCCTGGACAGCAGATTTACCGCATGCT 1345
1132 CAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1191
1346 TTTTGGTACGAGGCTGGGCTGCTGAGTGAACCTGCTGCTGCTGCTGCTGCTGCTGCTG 1405
1192 TTCTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1251
1406 TTTTGGGCAACCCGACCTGTAACCAACACCTTTTGGTTTC----- 1447
1252 TTGGCAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1311
1448 -----CAGACGATGTTTCAAGACGGGTACATATGCTGCTGCTGCTGCTGCTGCTGCTG 1498
1312 CCGGAGTGGGAGCGGGGTGTTACAGACATTTTGAAGTCCACAGCCGATTCATGTC 1371
1499 CGACCAGACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1558
1372 AAGCGGACCAACCCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1431
1559 GCGACCAACTTGGGAGGTTATTTGGCCAGCTCCCGGACATTTATGATTCAGGCGCTT 1618
1432 GCGACATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1491
1619 GCATTCATCTTGGGAGGCTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1678
1492 GTACTGACATACCGGAGTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1551
1679 AACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1738
1552 CCGACATCAAGGTTATATATCTTCAACAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1611
1739 ACGGCTTGGGAGAACCCGACACTTACGTGATGTGCGACGCTATCAAGACGCGAGG 1798
1612 GATGCGAGTGGCTGCAACCCGACTGCTGAGAGCGTGGGTATGCGCTCAAGGCTGAG 1671
1799 AGCATCATCAATACACT 1815
1672 GCTGATGCAATTAATCT 1688

RESULT 10
LOCUS E33665 1946 bp DNA linear PAT 31-JAN-2002
DEFINITION Novel protein, gene encoding the same and method for utilization
thereof.
ACCESSION E33665
VERSION E33665.1 GI:18624149
KEYWORDS JP 2000083675-A/1.
SOURCE Tricholoma matsutake
ORGANISM Tricholoma matsutake
REFERENCE Agaricales; Tricholomataceae; Tricholoma.
AUTHORS Takakura,Y., Kuwata,S. and Ota,S.

TITLE Novel protein, gene encoding the same and method for utilization
JOURNAL Patent: JP 2000083675-A 1 28-MAR-2000;
JAPAN TOBACCO INC., SOCIETY FOR TECHNO-INNOVATION OF AGRICULTURE
FORESTRY AND FISHERIES
COMMENT OS Tricholoma matsutake
PN JP 2000083675-A/1
PD 28-MAR-2000
PE 08-SEP-1998 JP 1998270606
PR YOSHIMITSU TAKAKURA, SHIGERU KUMATA, SHOZO OHTA
PI PC C12N15/09,A01N37/12,A01N63/00,A01N63/02,A01N65/00,C07K1/14, PC
C07K1/18,
PC C07K1/30,C07K14/375,C12N1/19,C12N1/21,C12N5/10,C12N5/10, PC
C12P21/02,
PC C12Q1/68//C12N15/09,C12R1:645),C12N15/00,C12N5/00,C12N5/00,
PC (C12N15/00,C12R1:645)
CC
FH Key Location/Qualifiers
FT source 1. 1946
FT /organism="Tricholoma matsutake".
FEATURES
source 1. 1946
Location/Qualifiers
1. 1946
/organism="Tricholoma matsutake"
/mol_type="genomic DNA"
/db_xref="taxon:40145"
BASE COUNT 484 a 516 c 475 g 471 t
ORIGIN
Query Match 6.1%; Score 129; DB 6; Length 1946;
Best Local Similarity 52.5%; Pred. No. 3.3e-16;
Matches 387; Conservative 0; Mismatches 305; Indels 45; Gaps 3;
QY 1106 CTGGGAGGATGCGACGAGCGGATGTCATTTGGCAGATGCTTGAGCGGAA 1165
DB 1064 CTTGGTCGCTACTCTCAGCAGACGTCATGACTTTTGTGATGTTCTCAAGAGAGC 1123
QY 1166 TTGCTGACAGCGCTGGCGCAGATCCTTATGAGACTGCATGTGGAAGAAGCGTTGCT 1225
DB 1124 ATAGTGATGATTCATCGCTACTGACCC-----CGCTTGGCTGGGAAGGTTGAG 1171
QY 1226 CAACATATTGCCAAGAACCCGACAGATGCATGCCATTCGTTCCGCGATCCGGAACC 1285
DB 1172 GCGCACAAGAGAAGAGACCCGATGAGCTGCTCCGATTCATTCACAGAGCTGAACCT 1231
QY 1286 CAGGTAAACAACCCATTTACAGAGAACAACCCCTGGACACGAGATTCACCGCGATGCT 1345
DB 1232 CAGGTATGATTTCTGACAGCTGCGACTTCCCTTGGATGTTCAAGGTCATTCGATGCA 1291
QY 1346 TTTTCGTACGTCGCTCGTCTCTGAGTGAGCTCTCGTCATGCTGACCTGCGCTGG 1405
DB 1292 TTT-----GCTGATTTGGAGCCCAAGGCCGACCCGCTGTGTCGATCTGAGGTTT 1345
QY 1406 TTTGGCGCAACGACCCCTGAAGCAACCTTTTGGTTTCC----- 1448
DB 1346 TTGCGCAATTCAGATTTGTCGAAGAATAATCGAGTACCTTCGTCGGAACCTTAAGCTA 1405
QY 1449 -----AGAACATGTTCAAGCGGGTACAGATGCGGAGCGAGCGATCAAGATAT 1498
DB 1406 CCGGACTGGGAAGCGGGTGTATACAGACTTATGATGACCAAGCGGATTCATGTC 1465
QY 1499 CGAACCCAGACCTGCTCAAACTGAGAGCAAGGAATAATGATGCGATATGTCGAAGTG 1558
DB 1466 AAGCGGACCAAGCCGATGGAAGCGTGAACAGAGATGATGATGATGATGACCAAGCTC 1525
QY 1559 GCGAGCAACTGGGAGGTTATTTGGCCAGCTCCCCCGGAGTTTATGATCCAGCGCTT 1618
DB 1526 GCGAATATCTGCGCGGATACCTTCTGCTCTACCTCAATTTATGGAACCTGCTGCTC 1585
QY 1619 GCACCTCATCTGGGGACTACTGCGATGCTTGGACAAAGCAACTCAAGGGGCTGAT 1678
DB 1586 GCACAGCATATCAGGGAAGCTACTCGATCGGACAGATGATCAAACTTGTGTGCTGAT 1645
QY 1679 AACCACTCGCTGCTGGGACTTTGCCAATCTTATGTTGACGAATGCGACCAATCAGG 1738

DB 1646 CCAGACATCAAAAGGTTACATCTTCGACAAATCTGGGCTGGCGGAATGGGTCAATCCA 1705
QY 1739 ACGGCGCTGGCGGACGAACCCGACACTTACTGATGTGGCCACCTATCAAGACCGGAGG 1798
DB 1706 GATGCGACTGCTGCAACCCGACTCGTACAGCGCTCGATGCTTAAAGGCTGTGAG 1765
QY 1799 AGCATCATCAATACACT 1815
DB 1766 GCTGTAGTCACTTACT 1782
RESULT 11
E37379 1946 bp DNA linear PAT 31-JAN-2002
LOCUS E37379
DEFINITION Reagent for assaying pyranose containing pyranose oxidase.
ACCESSION E37379
VERSION E37379.1 GI:18626680
KEYWORDS JP 2000175698-A/1.
SOURCE Tricholoma matsutake
ORGANISM Tricholoma matsutake
Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
Agaricales; Tricholomataceae; Tricholoma.
REFERENCE 1 (bases 1 to 1946)
AUTHORS Takakura,Y., Kuwata,S., Ota,S. and Usami,S.
TITLE Reagent for assaying pyranose containing pyranose oxidase
JOURNAL Patent: JP 2000175698-A 1 27-JUN-2000;
JAPAN TOBACCO INC
OS Tricholoma matsutake (pine mushroom)
PN JP 2000175698-A/1
PD 27-JUN-2000
PE 16-DEC-1998 JP 1998357423
PR YOSHIMITSU TAKAKURA, SHIGERU KUMATA, SHOZO OHTA, SATORU USAMI
PI PC C12Q1/26,C12N9/04,C12N15/09//G01N33/66,(C12N9/04,C12R1:645), PC
C12N15/09,C12R1:645),C12N15/00,(C12N15/00,C12R1:645) CC
FH Key Location/Qualifiers
FT source 1. 1946
FT /organism="Tricholoma matsutake (pine mushroom)"
FEATURES
source 1. 1946
Location/Qualifiers
1. 1946
/organism="Tricholoma matsutake"
/mol_type="genomic DNA"
/db_xref="taxon:40145"
BASE COUNT 484 a 516 c 475 g 471 t
ORIGIN
Query Match 6.1%; Score 129; DB 6; Length 1946;
Best Local Similarity 52.5%; Pred. No. 3.3e-16;
Matches 387; Conservative 0; Mismatches 305; Indels 45; Gaps 3;
QY 1106 CTGGGAGGATGCGACGAGCGGATGTCATTTGGCAGATGCTTGAGCGGAA 1165
DB 1064 CTTGGTCGCTACTCTCAGCAGACGTCATGACTTTTGTGATGTTCTCAAGAGAGC 1123
QY 1166 TTGCTGACAGCGCTGGCGCAGATCCTTATGAGACTGCATGTGGAAGAAGCGTTGCT 1225
DB 1124 ATAGTGATGATTCATCGCTACTGACCC-----CGCTTGGCTGGGAAGGTTGAG 1171
QY 1226 CAACATATTGCCAAGAACCCGACAGATGCATGCCATTCGTTCCGCGATCCGGAACC 1285
DB 1172 GCGCACAAGAGAAGAGACCCGATGAGCTGCTCCGATTCATTCACAGAGCTGAACCT 1231
QY 1286 CAGGTAAACAACCCATTTACAGAGAACAACCCCTGGACACGAGATTCACCGGATGCT 1345
DB 1232 CAGGTATGATTTCTGACAGCTGCGACTTCCCTTGGATGTTCAAGGTCATTCGATGCA 1291
QY 1346 TTTTCGTACGTCGCTCGTCTCTGAGTGAGCTCTCGTCATGCTGACCTGCGCTGG 1405
DB 1292 TTT-----GCTGATTTGGAGCCCAAGGCCGACCCGCTGTGTCGATCTGAGGTTT 1345
QY 1406 TTTGGCGCAACGACCCCTGAAGCAACCTTTTGGTTTCC----- 1448

Db 1346 TTGGCAATGATGATTTGCGAAGAAATGAGTGTCTCGTCCGAACCTTAAGCTA 1405
OY 1449 -----AGAACGATGTTCAAGACGGGTACAGTATGCGCAGCCAGCTTCAGATAT 1498
Db 1406 CGGCACCTGGGAAGCGGGTGTACAGACACTTATGSAATGCCAGCCACATTCATGCTC 1465
OY 1499 CGACCCAGCAGCTGCTCAACGCTGAGAGCAAGAAATGATGCGCATATGTGCGAAGTG 1558
Db 1466 AAGCGGACCAACCGCATGAGACCGTGACCAAGAGATGATGAATGATGACCAACGTC 1525
OY 1559 GCGAGCACTGGGAGGTATTTGGCCACGTCGCCCGCGAGTTATGATCCAGCCCTT 1618
Db 1526 GCGAACAATGAGGGGGGTACCTCTGCTCTCACTCAATTTATGCGACCTGGTCTC 1585
OY 1619 GCACCTCATCTTGGGGGAGTACTGCAATGCTTGCAGAAAGCACTACAGTGGCTAT 1678
Db 1586 GCACAGCAGATCAGCGGAGAACTACTGCGAGACAGATGATCAAACTTCCTGTTGCTAT 1645
OY 1679 AACCACTGCTGCTGTGGGACTTTGCCAATCTTTATGTTGCAAGCAATGCAACATCAGG 1738
Db 1646 CCGAATCAAAAGTTCATACCTGACATCTGTGGGTGCGGCGGAATGGGTGCATTCGA 1705
OY 1739 ACGGCTTCGGGAGAACCGGACCTTACGTGATGTCGACGCTATCAAGAGCGGAGG 1798
Db 1706 GATCGAGCTGCTTCGACACCGACTGTCAGAGCGTGCCTATGCGCTTAAGGGTCTGAG 1765
OY 1799 AGCATCATCAATACACT 1815
Db 1766 GCTGTAGTCACTTACCT 1782

RESULT 12
AB043883 1946 bp mRNA linear PLN 29-NOV-2002
LOCUS Tricholoma matsutake p20 mRNA for pyranose 2-oxidase, complete cds.
DEFINITION AB043883
ACCESSION AB043883
VERSION AB043883.1 GI:25553432
KEYWORDS
SOURCE Tricholoma matsutake
ORGANISM Tricholoma matsutake
Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes; Agaricales; Tricholomataceae; Tricholoma.

REFERENCE
AUTHORS 1 Takakura, Y.
TITLE Purification, characterization and primary structure of a pyranose
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1946)
AUTHORS Takakura, Y.
TITLE Direct Submission
JOURNAL Submitted (26-MAY-2000) Yoshimitsu Takakura, JT Plant Innovation
Center; 700 Higashibara Toyoda-cho, Iwata-gun, Shizuoka-ken
438-0802, Japan (E-mail: yoshimitsu.takakura@ims.jti.co.jp,
Tel: 81-538-33-8291, Fax: 81-538-33-6046)
FEATURES
Source Location/Organism
1. 1946 /organism="Tricholoma matsutake"
/mol_type="mRNA"
/db_xref="taxon:40145"
/note="synonym: Tricholoma nauseosum"
1. 1946 /gene="p20"
101. 1795 /gene="p20"
/EC_number="1.1.3.10"
/codon_start=1
/product="pyranose 2-oxidase"
/protein_id="BAC24805.1"
/db_xref="GI:25553433"
/translation="MPRLSKREKINDLQROGDLTSSODEIVHYTDFVLAGSGPIAC
TYAHIIDNTSTKVVMAETGSDONPYVIAHHRSTKRFQDTEKFEVNLINCAOPISIT
SPSDTYQPTLAVAMAPPIDBAEQVLYIMGNPNQEGNLNPGSAVTRIVGGMATHTMT

CACPTPDEERNVNEVDKQEFDALIERAKTLLNHSDOYDLSHQIYVKEFTLQOTLDA
SRGVTTLPLGVERRTDNP IYVWTGADVLVDVPSRPERVLTETRVKFI VSETNPT
OVVAALLRNLTNSIDELVVAOSFVJACAVATPOLILMSNLRPHALGRTYLSBOGMTFC
OIVLKRSIVDSIADPREFAKVEAHKHKHPDDVLP IPEHEPOVMILYSTDPMHVO
VHRAFKGVSFKADPRVVDLRFEGKSIVDENRYETGPNKLDMEWGVTDTGMPO
PPIHYKRTNAGDNDQRMNMTNANLTGSLYGLSGYQFAPARGLAQHTTTRIGTD
DOTSVADPTSKVHNEDNLVGNNGCTIPATCNPRISVAVALGAEAVSYLGV5-
BASE COUNT 484 a 516 c 475 g 471 t
ORIGIN
Query Match 6.1%; Score 129; DB 8; Length 1946;
Best Local Similarity 52.5%; Pred No. 3.3e-16;
Matches 387; Conservative 0; Mismatches 305; Indels 45; Gaps 3;
OY 1106 CTGGGAGGTACATCAGCAGACCGGATGCGATTTTCCAGATAGTCTTAGGCGAGAA 1165
Db 1064 CTGGGTGCTTACCTCAGGCAACAGTCATCTTTTGTCAAGATGTTCTCAAGAGAGC 1123
OY 1166 TTCTGCGACAGCGTGGCGAGAGATCTTATGAGACTGCGATGGAAGAACCGTTCCT 1225
Db 1124 ATAGTCGATTCATGCTACTACCT-----CGCTTCGCTGCGAAGGTTGAG 1171
OY 1226 CAACATATTTGCCAAGAACCCGACAGATGACATCGCCATTCGTCGCGATCCGGAACC 1285
Db 1172 GCGCAGAGAAGAACGCCCGGATGACGTGTCGCGATTCATTCACGAGCCTGAACTT 1231
OY 1286 CAGGTAAACACCCATTTTACAGAAACACCCCTTGCGACAGCAGATTCACCGGATGCT 1345
Db 1232 CAAGTGATGATTCCTGATACCGTGGACCTTCCTTGCGATGTCAGAGCTCATCGGTATGCA 1291
OY 1346 TTTTCGTACGGTGGCGGTGCTGTAGGTGACATCTGCTGTATCTGTGACCTGCGCTGG 1405
Db 1292 TTT-----GGTGTATTTGGACCCAGGCGACCGCGTGTGTGCTGTGATCTGAGGTTT 1345
OY 1406 TTTGGCGCAACCGACCCCTGAAGCAAAACCTTTTGTTGTTTCC----- 1448
Db 1346 TTGGCAATATCAGATATTTGTGAAAGAAATGAGAGTCTTTCGGTCCGAACCTTAAGCTA 1405
OY 1449 -----AGAACGATGTTCAAGACGGGTACAGATATGCGCAGCGGCTTCAGATAT 1498
Db 1406 CGCGACTGGAAAGGGGTGTACAGACACTTATGGAATGCCACAGCCGACATTCATGTC 1465
OY 1499 CGACCCAGCAGCTGCTCAAAAGCTGAGAGCAAGAAATGATGCGCATATGTGCGAAGTG 1558
Db 1466 AAGCGGACCAACCGCATGAGACCGTGACAGAGAGATGATATACCAACGTC 1525
OY 1559 GCGAGCACTGGGAGGTATTTGGCCACGTCGCCCGCGAGTTTATGATCCAGGCGCTT 1618
Db 1526 GCGAACAATGAGGGGGGTACCTTCCTGCTCTCACTCAATTTATGACACCTGTGCTC 1585
OY 1619 GCACCTCATCTTGGGGGAGTACTGCGATGCTTGCAGAAAGCACTACAGTGGCTGAT 1678
Db 1586 GCACAGCAGATCAGCGGAGAACTTGTGATCGGACAGATGATCAAACTTCTGTGTGAT 1645
OY 1679 AACCACTGCTGCTGTGGGACTTTGCCAATCTTTATGTTGCAAGCAATGCAACATCAGG 1738
Db 1646 CCGAATCAAAAGTTCATACCTGACATCTGTGGGTGCGGCGGAATGGGTGATTCGA 1705
OY 1739 ACGGCTTCGGGAGAACCGGACACTTACGTGATGTCGACGCTATCAAGAGCGGAGG 1798
Db 1706 GATCGAGCTGCTTCGAACCCGACTGTCAGAGCTGCGGTATGCGCTTAAGGGTCTGAG 1765
OY 1799 AGCATCATCAATACACT 1815
Db 1766 GCTGTAGTCACTTACCT 1782

RESULT 13
AB070201 2013 bp mRNA linear PRI 16-AUG-2001
LOCUS Macaca fascicularis testis cDNA clone: Qtsa-17226, full insert
DEFINITION
ACCESSION AB070201

[illegible]

<hr/>					
BC049706	RESULT 14				
LOCUS	BC049706	589 bp	mRNA	linear	ROD 01-APR-2003
DEFINITION	Mus musculus, clone IMAGE:6708084,		mRNA.		
ACCESSION	BC049706				
VERSION	BC049706.1	GI:29436895			
KEYWORDS					
SOURCE	Mus musculus (house mouse)				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
REFERENCE	1 (bases 1 to 589)				
AUTHORS	Strausberg,R.				
TITLE	Direct Submission				
JOURNAL	Submitted (31-MAR-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA				
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov				
COMMENT	Contact: MGC help desk Email: gcgbs@mail.nih.gov Tissue Procurement: Dr. Jonathan Kuo, NIMH CDNA Library Preparation: Michael Brownstein Laboratory CDNA library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Genome Sequence Centre, BC Cancer Agency, Vancouver, BC, Canada info@cgsc.bc.ca Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield, Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guln, Lettitia Hsiao, Martin Krzywinski, Reta Kusche, Oliver Lee, Soo Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven Ness, Pawan Pandoh, Anna-Lisa Prabhu, Parvaneh Saeedi, Jacqueline Schein, Duane Smalls, Michael Smith, Lorraine Spence, Jeff Stoltz, Michael Thorne, Miranada Tsai, Natasha van den Bosch, Jill Vardy, George Yang, Scott Zuydamduyn, Marco Marra. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Series: IRAL Plate: 45 Row: n Column: 24 This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis. Location/Qualifiers 1..589 /organism="Mus musculus" /mol_type="mRNA" /db_xref="taxon:10090" /clone="IMAGE:6708084" /tissue_type="Testicle, mouse" /clone_id="NH_MGC_169" /lab_host="DH10B" /note="Vector: pDNR-LIB"				
BASE COUNT	314 a	96 c	88 g	91 t	
ORIGIN					
Query Match	4.5%;	Score 94.4;	DB 10;	Length 589;	
Best Local Similarity	80.9%;	Pred. No. 3.9e-09;			
Matches 110;	Conservative 0;	Mismatches 26;	Indels 0;	Gaps 0;	
Dy	1971 TGCTGATTAAGCAGACTGTACTTTGGTCGATAGCTTAGGCGACTGATGCATTGATCATGTTTAAA	20300			
Dd	375 TGATTATTATTATTTCCCGAGATTAAATGATGAATAATTAAAGCCATATGTGTTCAAAAAAAAAA	434			
Dy	2031 AA	20900			
Dd	435 AA	494			
Dy	2091 AAAAAAAAAAAAAAAAAAAAAA 2106				
Dd	495 AAAAAAAAAAAAAAAAAAAAAA 510				

